Performance of CDMA Power Control and Admission Control in Multi-Service Cellular Systems

by

Wei Huang

B. Eng., Zhejiang University, Hangzhou, PRC, 1984
M. Eng., Southwest Jiaotong University, Chengdu, PRC, 1990

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Electrical and Computer Engineering

We accept this dissertation as conforming to the required standard

---

Dr. Vijay K. Bhargava, Supervisor (Department of Elect. & Comp. Eng.)

Dr. Wu-Sheng Lu, Member (Department of Elect. & Comp. Eng.)

Dr. Panajotis Agathoklis, Member (Department of Elect. & Comp. Eng.)

Dr. Hans A. Muller, Outside Member (Department of Computer Sciences)

Dr. Cyril S. Leung, External Examiner (Department of Elect. & Comp. Eng. UBC)

Supervisor: Dr. Vijay K. Bhargava

Abstract

This dissertation is focused on multi-service and direct sequence code division multiple access (DS/CDMA) wireless cellular systems.

The reverse link performance of a slotted DS/CDMA cellular system with multi-service traffic is analyzed. Services with/without packet retransmission to meet their Quality of Service, share the entire bandwidth. Packet failure probabilities and packet delay are obtained based on analyzing the mutual interaction among services. The impacts of power level allocation and power control error of services on capacity, throughput and delay are analyzed under given Quality of Service. The system capacity is maximized by appropriate power allocation. The impact of power control error on capacity is dependent on whether packet retransmission is allowed or not.

Admission control policies for multi-service systems are proposed and analyzed. Both nonprioritized and prioritized admission control are studied. Services difference in terms of resource requirement and degree of importance are considered. Analytical models are developed. Blocking probability of each type of calls are found under given amount of traffic. Fair access by soft capacity is addressed. The cost of protecting certain type(s) of calls on the rest of calls is investigated. The impact of traffic distribution on the performance of the policies is also examined.

In a hierarchical cellular system, user mobility estimation helps channel assignment so as to reduce the handoff rate and avoid high mobility users travel among small cells. Two different strategies are compared. It is found that when high mobility users are served by overlay macrocells, call drop rate is reduced. Speed estimation error only has limited impact on the system performance. User membership in a cellular CDMA network is simulated based on the estimation of the local mean value of the pilot signal from surrounding base stations. The base station providing strongest pilot local mean controls the mobile station. Simulation is conducted under different fading environments. Two performance measurements are simulated: the number of membership switchings per second and the probability of wrong base station selection. An optimum window length for filtering out Rayleigh fading is found. Simulation results are in good fit with those of analysis.

Examiners:

Dr. Vijay K. Bhargava, Supervisor (Department of Elect. & Comp. Eng.)

Dr. Wu-Sheng Lu, Member (Department of Elect. & Comp. Eng.)

Dr. Panajotis Agathoklis, Member (Department of Elect. & Comp. Eng.)

Dr. Hans A. Muller, Outside Member (Department of Computer Sciences)

Dr. Cyril S. Leung, External Examiner (Department of Elect. & Comp. Eng. UBC)

Table of Contents

List of Tables

List of Figures

Acknowledgments

First, I would like to thank my supervisor, Dr. Vijay K. Bhargava for giving me this opportunity to pursue my Ph. D. program in the University of Victoria. Without his guidance, encouragement and support, this research would not have been possible.

I would also like to thank Drs. Fortunato Santucci, Roman Pichna, Qiang Wang, Jialin Zou and Bo Wu for many inspiring discussions and advice. Thanks to my colleagues in the Communication Research Lab for being helpful, cooperative and supportive. Special thanks to Mr. A. Annamalai for his support during my thesis revision while I am away from the University of Victoria.

My gratitude extends to my committee members for their evaluating my thesis.

This research is funded in part by the Canadian Institute of Telecommunications Research (CITR) and by the Natural Sciences and Engineering Research Council (NSERC) of Canada under a strategic project grant. Nortel has also provided research fund.

Finally, my wholehearted appreciation goes to my family members. Their encouragements, sacrifice and unlimited love are the driving force of me to accomplish this program and to seek a bright future.

To my parents

Huang Maoxi and Miao Shucheng

and my wife

Jiang Yawen

Chapter 1
Introduction

1. 1 Motivation of Research

In recent years, wireless communications become the fastest growing area of telecommunications. It attracts the attention of researchers, industry and consumers. Tremendous efforts has already been invested into this area in order to develop more advanced systems with higher capacity, more services and lower costs. The time that people are freed from tethers has already been envisioned. Since frequency spectrum is the most precious resource for the running of wireless systems, all users are required to share this resource efficiently by the so called multiple access schemes. Signals of different transmissions can be separated at receivers by using any of those schemes or their combination and the interference among different signals can be controlled. There are three most common schemes: frequency division multiple access (FDMA), time division multiple access (TDMA) and code division multiple access (CDMA). A component of the entire resource domain, such as a frequency band, a time slot or a code, is assigned to a communication link.

The evolution of wireless communication systems is usually divided into three phases [1]. During 1980', the first phase, the first generation systems were put into market. The advanced mobile phone service (AMPS) was a typical one of them. For this generation, analog technology was used which provided limited features. The cellular concept was adopted as well to achieve frequency reuse. The multiple access scheme used was FDMA resulting in rather small system capacities. The second generation systems use digital technology to improve their performances over the first generation. Most of them are TDMA ones, while Qualcomm put forward a CDMA solution [27]. CDMA is well known for its capability of resolving multipath and interference suppression. Based on extensive research and development in recent years, CDMA cellular phone systems have begun to be deployed in a number of countries and to compete with their TDMA counterparts. The second generation systems mainly only support voice service. The third generation systems will be designed to be capable of efficiently handling heterogeneous traffic [45]. More and more services originally supported by wireline systems are going to be carried by the third generation systems. The third generation systems, to be launched around the year 2000, face huge capacity demand, enhanced service quality requirements, and efficient service integration. The earlier ambitious goal of the future cellular systems, which is to deliver information on an "any time, any where, any one" basis, has been expanded to include "any media".

Before the third generation systems can be deployed, there is still a lot of research and development work remaining. We have to address many new challenges introduced by the coexistence of multi-service. Admission control is one of them, which greatly affects the handoff performance of a system as well. In a multi-service scenario, the service dissimilarity, the mutual influences due to resource sharing, as well as our design targets will affect the final choice of admission control policy. Another closely related factor involved is to enhance handoff performance in order to reduce forced call terminations, handoff signalling traffic load, and signal degradation. Handoff is an important issue in any mobile cellular system. Handoff becomes more frequent due to the cell size shrinking for a higher capacity and the support of higher mobility users.

CDMA technology is a candidate of the third generation system multiple access scheme and under intensive investigation [25], [47]. In Canada, an integrated wireless access network (IWAN), a multi-service CDMA cellular network, has been proposed and studied [26], [48]. In Europe, a framework for third generation CDMA system was also carried out under the CODIT project [2]. Power control and power allocation are essential to the operation of multi-service CDMA networks. Power control is a mechanism to keep the power level at all receivers staying at a preferred level and power allocation determines what is the preferred level of each service. In our research, both multi-service and CDMA related issues are chosen. The results provided will be valuable for further research and development on the third generation systems.

1. 2 Contributions of the Dissertation

The objectives of this research are: to investigate the impact of power control on direct sequence CDMA (DS/CDMA) system capacity and propose power control law for multi-service systems; to propose and analyze multi-service admission control policies; and to study strategies to improve handoff performances.

The reverse link capacity of a multi-service CDMA system is analyzed. An analytical model is developed to calculate the system capacity and its relationship with power allocation and power control error. Two scenarios are considered: system carrying voice and data services and system carrying two different data services. Results show that proper power allocation can maximize the capacity. Power control errors affect the system capacity, packet delay, and throughput. Power control laws are proposed.

The impacts of service coexistence and dissimilarity on the admission control policy (ACP) design are also addressed. Admission decisions are made based on certain call parameters, including the degree of importance and the required amount of resource of a call, as well as the admission rule. We developed two-dimensional Markov chain models to obtain the performance of different ACPs in terms of blocking probability and fairness indicator. Certain types of calls can be protected either by its access enhancement or by others access limitation. The cost of call protection is given. The proposed ACPs cover a variety of control requirements.

We study handoff performance in a hierarchical cellular system with user mobility considered. Efforts are made to reduce the handoff rate as well as handoff failure probability by properly assigning channels to mobile stations (MSs). In addition, the user membership detection, an integral part of handoff procedure, is also addressed. Simulations are conducted to find out the performance in terms of power control switching rate of an MS among neighboring base stations (BSs) and the probability that an MS is controlled by a BS such that higher interference is generated. The ways to improve the performance are addressed.

1. 3 Outline of the Dissertation

In Chapter 2, fundamentals of wireless cellular systems, CDMA technology, teletraffic and Markov chains are briefly presented. Previous works are highlighted.

Chapter 3 focuses on the power control and power level allocation issues in a multi-service DS/CDMA network. A slotted ALOHA based access protocol is assumed in the reverse link. Services with/without packet retransmission are considered. Two cases of two service system are investigated. The impact of power control error and power allocation on the capacity, delay and throughput is provided.

Chapter 4 deals with the admission control policies of multi-service systems. Different services may require different amount of system resource, such as power allocation or bandwidth, to carry a call. Admission control policies coping with this fact are analyzed by two-dimensional Markov chain models. The impact of protecting important services/calls by means of channel allocation is addressed. Both the impact of traffic load distribution on admission performances and the fairness factors are given.

In Chapter 5, the impact of user mobility and handoff strategy on the handoff performance is addressed. User membership switching in a CDMA system is also simulated to find the handoff performances and to justify analytical results.

In Chapter 6, both concluding remarks and suggestions for further research are presented.

Chapter 2

Fundamentals and Previous Works

2. 1 Cellular and CDMA Basics

In this chapter, some useful background information, concepts, and terminologies are reviewed to help the explanation of the following chapters. In addition, previous works, within the areas covered in the following chapters, are also reviewed.

A wireless cellular system uses a lot of BSs to cover its service area. Each BS serves a geographical area, called a cell, which is usually represented by a hexagon. All MSs in a cell communicate with the BS in that cell. The transmission power level used by BS and MSs in a cell is limited only for in cell communications. Therefore, the spectrum can be reused simultaneously in different cells separated by enough space such that the cochannel interference is controlled to be lower than a certain level [6]. By frequency reuse, limited spectrum can serve nearly unlimited number of mobile users. If the cell size is small, MSs can transmit at low power level and thus have a long battery recharge cycle.

As digital technology matures, digital cellular systems have been put into market. These systems have a number of advantages over their analog counterparts. They are more flexible, easier to implement encryption, have more natural integration with digital wireline systems and easier to reduce the source data rate by signal processing. Digital technology is the base of present TDMA or CDMA systems. It is also the base of TDMA or CDMA third generation systems proposed in [2], [25].

Spread spectrum is a communication technique originally used in the military. It is proposed for commercial use as a multiple access protocol. One of the CDMA schemes, DS/CDMA, is based on spread spectrum technology. In DS/CDMA communication systems, a unique high rate spreading sequence is assigned for each call to each user. By multiplying this sequence with the user data, the user's signal bandwidth is spread to a much wider bandwidth. Different sequences used in different links must have very low cross-correlation. At the receiver end, the individual user's signal can be separated from others with a correlator by despreading. The correlator in a receiver is synchronized to the received signal with the same spread sequence used in spreading. As shown in Fig. 2. 1, other signals are not despread and contribute to interference. In a DS/CDMA system, the processing gain is defined as:

$$PG = \frac{W_s}{W_d} = \frac{T_d}{T_s}, \tag{2.1}$$

where W_s and W_d are the bandwidth of the spread spectrum signal and bandwidth of user signal before spreading, respectively. T_d and T_s are the time duration of a data bit and the time duration of a spreading chip, respectively. The higher the processing gain, the higher the receiver's capability of interference mitigation. The system capacity defined as the number of active MSs in a cell is determined by a number of factors as [27]:

$$N \approx \frac{W_s/W_d}{E_b/N_o}\alpha F G, \tag{2.2}$$

where α, F, and G are voice activity factor, frequency reuse efficiency and the number of sectors in a cell, respectively.

The main reason of choosing CDMA for present and future systems is that CDMA can provide a higher system capacity for a given amount of spectrum. Universal frequency reuse is adopted in CDMA systems, but not in FDMA and TDMA systems. Universal frequency reuse removes the necessity of frequency planning and management, and makes soft handoff feasible. CDMA also has some other features considered being advantages over the traditional FDMA and TDMA [33]. Among them are:

- multipath mitigation capability: multipath signals with path delay difference larger than a chip duration can be resolved by a RAKE receiver; at BSs and MSs, RAKE receivers are used to resolve multipath signals;
- voice activity cycles: no transmission (and therefore, interference) when people are listening (also true for any services that have an activity factor less than 1);
- soft capacity: quality of receiving signals gracefully degrades as the number of ongoing calls increases beyond a threshold;
- soft handoff: due to universal frequency reuse, an MS can communicate with more than one BSs during handoff procedure in order to improve the receiving quality. On the forward link, signals are sent via several BSs simultaneously. The signals are combined by the MS receiver. On the reverse link, corresponding BSs receive the MSs' signal copies and send them to a common node to combine.

A major disadvantage of DS/CDMA systems is that they need much tighter power control compared with other access protocols. This fact introduces extra soft/hardware complexity and implementation cost. The system performance is sensitive to the power control errors.

The IS-95 standard proposed by QUALCOMM is for digital cellular telephone systems based on CDMA technology [28]. Next generation CDMA systems are under investigation. They will carry not only voice service but also multi-media services, provide flexible air interface, and handle multimedia traffic.

2. 2 Power Control and Power Allocation

Since all BSs/MSs share the same frequency spectrum of the forward (BS to MSs)/ reverse (MSs to BS) links in DS/CDMA cellular networks, the system capacity is interference limited [20]. Power control is essential to the operation of DS/CDMA networks. It is a mechanism to keep the power level received staying at a preferred level in order to maximize the system capacity. In the reverse link without power control, an MS closer to a BS could be received at a higher power level than other MSs. If the difference between the received power levels is too high, the closer one may cause a great interference to other MSs. This is the so called near-far effect. Fortunately, the near-far effect on the forward links is not significant. In the forward link, interference comes from a few non-movable BSs in neighboring cells; in the reverse link, interference comes from a large number of movable MSs within the reference cell and within its neighboring cell as well. The fluctuation of reverse link interference is higher than that of the forward link. Due to the different interference nature of the forward and the reverse links, different power control schemes are used.

2. 2. 1 Forward link power control

In the forward link, it is desirable to provide a means of controlling the power received by an MS. A BS's transmission power to the MS is determined by the requests from the MS. The reason for introducing this type of control is to improve the receiving quality of MSs when forward channels become poorer and/or interference becomes higher. For example, an MS closer to the cell boundary may suffer higher intercell interference and higher path loss than the ones closer to a BS. Thus it is necessary for the BS to increase its power above its average power.

There are two schemes for forward link power control: distance-based and quality-based. In the first scheme, the transmitting power of a BS is a function of r, the MS - BS distance normalized by R (R is the cell radius) [33], [52]. The proposed power control law is in the form of $\varphi(r) = r^n$, where n is a constant and $r_t \le r \le 1$. For MSs with $r \le r_t$, the BS's power is fixed to $\varphi(r) = r_t^n$. For the rest of MSs in the cell, the BS's power increases as r increases. Therefore, BS's power is a function of the MS - BS distance. The author of [33] found the best values: $n = 2$ and $r_t = 0.55$. A certain signal to interference ratio (SIR) can be kept in the entire cell by this power control law. The same law was further improved by balancing the forward link power shared by all MSs in a cell and considering the intercell interference experienced by MSs closer to a BS [52]. Two set of the best parameters are reported: $n = 2$, $r_t = 0.6$ and $n = 3$, $r_t = 0.75$. The forward link capacity can be 200% (about 178% in [33]) of that without power control.

In the second scheme, however, the BS's power is adjusted according to the MS's receiving signal quality indicated by bit error rate (BER), SIR etc. MSs report their receiving quality regularly to their serving BS. The BS adjusts its transmitting power accordingly making the signal quality is just above threshold during most of the time, say 99%.

By comparing the above two schemes, it is found that the distance-based scheme is suitable for the case of no or less shadow fading environments. In that case, path loss, the only factor that should be compensated by power control, is proportional to r^n. However, the distance-based scheme does not work well in a shadow fading environment but a quality-based scheme does [37]. Since fading is a common phenomenon in cellular environment, we can not avoid it. In addition, the papers on distance-based scheme have not given the details about how to obtain r accurately. One of the possible ways of estimating r is using pilot strength measurement by the MS. As the distance estimations might have some error, performance of the distance-based scheme degrades. Therefore, distance-based scheme only provides a theoretical picture of how the average BS transmitting power level distributes within a cell. Quality-based scheme is more realistic and can cope with cellular environments. In the IS - 95 standard, quality-based approach is adopted. An MS reports its signal quality statistics to a BS either periodically or only upon the MS finds that the frame error rate of the forward link is higher than a threshold.

The dynamic range of forward link power control is smaller (6 dB) and its power control command rate is slower (once per 15 ~ 20 milliseconds), compared to those of the reverse link power control. If the dynamic range is too small, say 4 dB, the forward link capacity will be reduced by 10% under a given outage probability [39]. Forward link power control is very important as well. Without it, the forward link capacity will be greatly reduced, even lower than the power controlled reverse link [4]. Forward link power control is less addressed than its reverse link counterpart.

In the forward link, each BS transmits a pilot signal. On one hand, the capacity available for traffic reduces since the pilot interferes traffic channels. On the other hand, the pilot makes synchronous communications in the forward link possible and thus reduces the

E_b/N_o threshold. The overall impact of the pilot on forward link capacity is positive. The pilot power is controlled to be 20% of the total BS transmission power, higher than any other traffic channel power level, and easy to be tracked by MSs.

2. 2. 2 Reverse link power control

In the reverse link, the so called near-far effect exists. Power control aims to maximize the reverse link capacity. Since this capacity is less than the forward link capacity, power control in the reverse link is of our interest. There are two power control schemes: strength-based and SIR based (i. e. the target value of the controlled power is represented in terms of power strength and SIR, respectively). The strength/SIR-based scheme tries to keep the received power strength/SIR from all MSs equal and just above threshold. It is found that the strength based scheme is more stable but with a higher outage probability than the SIR-based power control [43]. For SIR-based scheme, an optimum power level exists, which is mainly a function of traffic load and difficult to obtain in real time. For strength-based scheme, the power level is fixed.

Other parameters affecting the performance are the order of power control command, the step size, dynamic range, BER of power control command, and the processing delay of the power control mechanism [69]. The *error* signal in Fig. 2. 2 must be quantized by means of pulse coded modulation (PCM has a power control command set of $\{-n, -(n-1), \ldots -1, 0, 1, \ldots n-1, n\}$) or delta modulation (DM has a command set of $\{-1, 1\}$). PCM outperforms DM by offering a lower outage rate. However, the improvement becomes insignificant for $n > 3$ [43]. Power control algorithm also affects performance. The performance of a variable step scheme is just a little bit better than that of a fixed step one. An algorithm based on fuzzy logic is proposed [42]. It has the advantages of faster rise time, less overshoot, and smaller root-mean-squared tracking error over the conventional algorithms. Since power control can not totally remove power fluctuation, the received power follows a log-normal distribution [69].

Power control in the reverse link can have two components [27]. One is the closed loop power control accomplishing fast convergence to the desired receiving level at the BS. Its model is shown in Fig. 2. 2. As the channel quality on both directions are usually not equal, BSs have to constantly observe the received signal strength and determine power control commands for the reverse link transmission. The second component is the open loop power control at the MSs, which provides a rapid response to a sudden improvement (but not a sudden degradation) in channel quality in order to eliminate excessive power level at the BS antenna. Based on the measurements on the forward link signal strength, MSs can set their transmitting power accordingly within a few microseconds.

It has been shown that a poor power controlled CDMA system dramatically loses its capacity compared with a well power controlled one [21], [34]. Table 2. 1 gives the reverse link capacity loss due to power control errors, which indicates that the power control error should be within 1 dB. Reverse link power control has a dynamic range of 80 dB. Smaller dynamic range, say 60 dB, will reduce the effectiveness of power control and reduce the system capacity by 39% [39]. The transmission rate of closed loop power control commands is 800 bps in the IS-95 standard. In CODIT project, 2 kbps power command is transmitted via a control channel. Although power command bits have BER, the power control performance is not sensitive to the BER [32], [63].

Table 2. 1: System capacity loss (%)

Power control error (dB)	Reference #		
	[37]	[38]	[4][a]
1	38	29	28 ~ 38
2	65	64	52 ~ 65
3	80	83	68 ~ 80
4	90	-	80 ~ 90

a. outage probability threshold: from 0.1 to 0.02.

2. 2. 3 Power allocation

Although the above concepts are first established for DS/CDMA wireless telephone systems, they can be extended and adopted under multimedia scenario by future CDMA networks. Unlike single traffic networks, a multimedia one has to carry a number of different services which share the allocated spectrum. These services have different quality of service (QoS) requirements usually in terms of data rate, BER threshold, and information delay threshold. For example, voice quality and video quality are sensitive to transmission delays. Data, email, and data base accesses usually can tolerate longer delays. The BER threshold of those services ranges from 10^{-3} to 10^{-9}. Their transmission rates can vary from a few kbps to 2 Mbps [25], [44]. A number of typical services are given in [45] along with their quality requirements in terms of BER, delay, and transmission rate. Due to those dissimilarities, each service should be assigned a suitable power level received at the BS to maximize the system capacity (There are also other possible ways, such as adaptive error control coding, to cope with the service dependent QoS requirements. But they are out of the scope of this work). It is called power allocation. All the MSs of a service are power controlled according to the power strength (or SIR) allocated to that service by the power control schemes discussed before. The power allocation of a particular service should be as low as possible to reduce interference on other receivers and high enough to assure its own QoS. Either strength based power allocation or SIR based power allocation are possible. We will discuss strength based power allocation in Chapter 3. As power control is already an integral part of CDMA systems, power allocation is a natural extension of power control and can be easily implemented compared to other means of dealing with the service dissimilarity.

CDMA networks with user data being spread over the entire allocated spectrum, regardless of the data's service type, are called single bandwidth systems. The IWAN system is one of them [26]. It has been shown that for such a system, its capacity is largely limited by service with high bit rate and high quality requirement. An optimum power allocation for different services, voice and video, is suggested, for the reverse link, under which all services have the same outage probability, in [35]:

$$P_i \approx \frac{PG_1 (E_b/N_o)_i}{PG_i (E_b/N_o)_1} P_1 = \frac{R_i (E_b/N_o)_i}{R_1 (E_b/N_o)_1} P_1, \tag{2.3}$$

where $(E_b/N_o)_i$ represents the QoS of type-i service with PG_i as its processing gain. For a single bandwidth system, we notice that processing gain is inversely proportional to the data rate, resulting in the right hand expression of Eq. (2.3). Therefore, power allocation is approximately proportional to service QoS and reversely proportional to service processing gain. It is also suggested that services with different data rate and the same QoS can be accommodated by allocating power levels proportional to their data rates [31]. In [55], reverse link power allocation is addressed under different system assumptions. Services are grouped into a low priority group and a high priority group. The idea is to adjust the power allocation of high and low priority services to keep the BER of the high priority ones below threshold. The goal of power allocation is to maximize the SIR of the low priority services while maintaining the BER of high priority ones. It is found that the BER of low priority services increases as the total traffic load increases. The proposed system outperforms the slotted ALOHA and TDMA systems in terms of the total throughput.

In multi bandwidth systems, such as CODIT, different services are spread over different bandwidths depending on their information data rates and their QoS requirements. Interfrequency handoff may be experienced, in this case, by MSs of narrow band services. During such handoffs, the power control mechanism doubles the allocated MS power level so that the MS enters the "compressed mode", which allows the MS to communicate with two BSs simultaneously to achieve a seamless handoff [25]. Therefore, the MS power is temporarily changed. By introducing the "compressed mode", dual radio MSs are unnecessary while the system can still support interfrequency seamless handoff. However, the performance of the "compressed mode" is not released. For multi bandwidth systems, power allocation rules have been obtained without considering the interfrequency handoff. Two power allocation rules, the equal error probability rule and the equal signal strength rule, are proposed and compared [56]. It is found that the equal error probability rule can provide a higher system capacity. Finally, since the services may use different QoS measures and/or different bit error control schemes, it is necessary to investigate how to allocate power among services with those differences.

2. 3 Teletraffic and Performance Measurements

2. 3. 1 Traffic model

For each cellular system, we must consider its traffic handling capability. The stream of access requests from MSs to a BS is a random process. Poisson processes are widely used to describe call arrival processes of cellular systems [9], [18], [41]. For such processes, we know that the time between the arrivals of two consecutive calls is negative-exponential distributed with mean $1/\lambda$. It is also assumed that the channel holding time is a negative-exponential distributed random variable with mean $1/\nu$. Therefore, the number of active calls in a cell is given by:

$$P(k) = \frac{(\lambda/\nu)^k}{k!} e^{-\lambda/\nu}. \tag{2.4}$$

Due to the limited cell size, MSs may cross cell boundaries during a call resulting in handoff traffic on top of new call traffic. During a call, an MS may spend successive periods of time (t_0, t_1, t_2...) in a number of cells as shown in Fig. 2. 3. At cell boundaries, link between MS and new/old BS has to be established/terminated by means of handoff. Handoff is an essential function to keep the service continuity in mobile cellular systems. Mobility generated handoff traffic affects the system performance. There are two kinds of handoff, hard handoff and soft handoff. With hard handoff, an MS can only communicate with one BS at a time; during soft handoff period, an MS communicates with more than one BS simultaneously. There are two phases in a handoff procedure.

In the first phase, handoff initiation phase, the need of handoff should be quickly and accurately identified. The major metrics to measure the handoff initiation performance are the probability of unnecessary handoff and the handoff delay. Since we can not minimize both metrics at the same time, trade-off has to be made. A number of schemes, as listed below, are proposed for the hard handoff initiation.

- Relative signal strength: An MS is handed off to a BS providing the strongest signal strength. This is the simplest scheme. It may introduce a lot of unnecessary handoffs due to the so-called ping-pong effect. In CDMA cellular systems, pilot signal strength measurements taken by MSs are used to determine the coverage of a cell and trigger handoffs. Since BS's signals suffer fast fading, its strength is not reflected by a single sample but mean value of a random variable. Therefore, a number of signal samples should be put through an average window to remove signal fluctuation due to fast fading. The averaging output is then used in handoff decision making. Of course, signal strength of an MS at a BS can also be used to trigger handoff. The choosing of a proper averaging window and the number of samples used is addressed in [11]. An alternative channel estimation method is based on the least squares estimation [57]. It outperforms long window averaging method in terms of having fewer handoffs to keep the same signal outage probability.

A threshold criteria can be added to this scheme. In this case, handoff initiates when the current signal is not the strongest one and below a threshold.

- Relative signal strength with hysteresis: To avoid most unnecessary handoffs, a hysteresis is introduced. In this case, handoff is initiated only when the current BS's signal is weaker than a new BS's signal by the hysteresis. Handoff initiation is thus delayed. Since large hysteresis increases handoff delay which may result in unacceptable receiving signal degradation and small hysteresis increases the number of unnecessary handoffs, the selection of a good hysteresis is crucial.

  A threshold criteria can be added to this scheme as well. To trigger a handoff in this case, the current signal strength has to be lower than a threshold and lower than a new signal by a given hysteresis.

In addition to the schemes based on signal strength measures, schemes based on carrier to interference ratio (CIR) are also proposed. In cellular systems, CIR is more reliable than signal strength as an indication of receiving quality because cochannel interference is taken into account [15]. Moreover, system may trigger handoffs to relief traffic overload in a cell.

For CDMA soft handoff, a BS-MS link is setup (terminated) when the BS's pilot signal strength is higher/lower than an adding (dropping) threshold. While an MS has links with more than one BS simultaneously, the MS is doing soft handoff. Therefore, a soft handoff is initiated when the second BS's signal strength becomes higher than the adding threshold. Soft handoff is used in CDMA cellular systems because the universal frequency reuse allows one radio operation at the MSs. Compared to hard handoff, soft handoff keeps the continuity of calls, extends cell coverage, introduces diversity combining in both directions to improve receiving quality, and reduces mutual interference. The disadvantage of soft handoff is that it requires more complex handoff control and system hardware overhead.

The role of MS and/or BS in handoff initialization can be quite different. In early systems, BS has the control of handoff initialization. In recent systems, however, MS can be used as an assistant to a BS or even takes charge of handoff triggering. If both BS and MS are involved in handoff procedures, handoff performance can be improved further. General reviews on handoff issues can be found in [15], [16] and [53].

In the second phase, the handoff execution phase, channel in the new cell must be assigned to handoff MS to finish the handoff. If there is no channel available, a handoff will be blocked or queued. Our focus in Chapter 4 is on the second phase performance.

It has been shown that handoff traffic can be recognized as Poissonian, which simplifies performance analysis [19]. The amount of handoff traffic depends on new call arrival rate λ_n, average call holding time, user mobility, sectorization, and the cell size. We denote the blocking probability of new call and the blocking probability of handoff as P_n and P_h, respectively. The handoff arrival rate λ_h is related to new call arrival rate as [14], [17]:

$$\lambda_h = \frac{\eta(1-P_n)}{\mu+\eta P_h}\lambda_n, \tag{2.5}$$

where $1/\mu$ is the mean of call duration time and $1/\eta$ is the mean of MS cell residence time. When P_n and P_h are far less than 1, handoff traffic is proportional to the new call traffic:

$$\lambda_h \approx \frac{\eta}{\mu}\lambda_n. \tag{2.6}$$

Due to handoff arrivals, the total call arrival rate in a cell becomes:

$$\lambda = \lambda_n + \lambda_h. \tag{2.7}$$

The channel holding time of MSs in a cell is a negative-exponential distributed random variable. However, because of handoff departures, the mean channel holding time in a cell is shorter than the average call duration time $1/\mu$:

$$\frac{1}{\nu} = \frac{1}{\mu + \eta}. \tag{2.8}$$

The traffic intensity ρ in a cell in Erlang is defined as:

$$\rho = \frac{\lambda}{\nu} \approx \frac{\lambda_n + \frac{\eta}{\mu}\lambda_n}{\mu + \eta} = \frac{\lambda_n}{\mu}. \tag{2.9}$$

Therefore, the overall result is that a nearly constant traffic load is observed in a cell [18]. Handoff traffic takes certain percentage of the total traffic ρ under given μ and η However, handoffs introduce handoff signalling traffic in both directions reducing the system capacity for data traffic. In addition, a higher handoff rate results in a handoff failure probability.

2. 3. 2 Performance measures

The cellular system performance measures widely used are the blocking probability of new call, the blocking probability of handoff call, and the probability of forced termination P_f [14], [18]. Blocking reflects the insufficiency of channels in a cell. Unfortunately, to eliminate blocking by deploying more channels in each cell can be costly or impossible. However, we can wisely use available channels to improve performance by means of admission control. Admission control is one of the aspects of cellular network management. It keeps the traffic load level in the system acceptable to meet the quality requirements of all users. It follows certain control rule to assign channels for the coming calls. For example, if the rule requires that new calls and handoff calls are treated identically and all blocked calls are cleared from the system, the performance will be given by the Erlang-B formula:

$$P_n = P_h = \frac{\rho^m / m!}{\sum_{i=0}^{m} \rho^i / i!}, \tag{2.10}$$

where m is the number of channels in a cell. When P_n and P_h are known, P_f can be obtained as in [17]:

$$P_f = 1 - \frac{1 - P_n}{1 + \frac{\eta}{\mu} P_h}. \tag{2.11}$$

Since handoff blockings are considered worse than new call blockings, it is desirable to reduce the handoff rate, which helps to reduce both handoff blocking probability and handoff signalling load at the same time. However, as the cell size shrinks to increase the system capacity, handoff rate is going to increase. The reduction of handoff blocking becomes even more important in order to keep the service quality satisfactory. Thus, different ways of giving handoff call priority are proposed for single traffic systems [49], [50]. Traditionally, handoff requests and new access requests are treated equally in channel assignment procedure. In this way, Eq. (2.10) gives P_h, which is usually considered being too high for handoffs. To reduce P_h, several priority schemes are proposed. In [9], a channel reservation scheme is addressed. A number of channels are reserved for handoff calls only. Therefore, new calls have fewer number of channels to access. It is not surprising that P_h is reduced while P_n increases as the number of the reserved channels increases. As user mobility increases, P_f increases faster if no channel is reserved for handoff.

Handoff queueing is another attractive method [58], [59]. It is based on the fact that there is overlap area between cells. Before an MS leaves the old cell, it has already entered a new cell. If the new cell can not assign the MS a channel, the communication is kept via the old BS and the handoff request is put into a queue rather than being blocked. Unless the MS really leaves the old cell's coverage, it can wait for the new BS to assign a channel to it, which helps reduce P_h. The queueing can be either based on a FCFS scheme (handoff request arriving the queue earliest is served first) or measurement-based prioritization scheme (handoff in the queue with the poorest signal quality is served first. Thus the overhead on signal quality monitoring is introduced) [59]. The latter outperforms the former in terms of having a lower P_h, a shorter average queue size and a shorter average waiting time. The difference on P_n between these two schemes is not significant.

If the transmission rate can be controlled, channel sub-rating helps to reduce P_h. The idea of sub-rating is to reduce the transmission rate of an existing connection by half such that a coming handoff can obtain a half rate channel for its temporary half rate transmission. In usual case, handoff will be blocked if there is no channel available in the target cell. Results show that P_h is substantially reduced at the cost of reduced service quality during half rate transmission [60]. Fortunately, the probability of a call being sub-rated is quite low (only 3% users experience 5 seconds or longer sub-rated conversation). The capacity is increased by 8 ~ 35% compared with the capacities of the other mentioned schemes. A limitation of this scheme is that some services may not be able to tolerate the quality degradation nor change their transmission rate easily.

The implementation of above proposals can effectively reduce the probability of no channel available for handoff calls such that we can reduce both P_h and P_f further [9]. However, the usual cost is a increased P_n compared with that in Eq. (2.10). Trade-off has to be made between new call performance and handoff call performance.

In multimedia systems, the incurred cost of a handoff blocking becomes service dependent. Therefore, the selection of admission policy will be affected by this fact in order to minimize the cost. Admission policies for single traffic scenario must be extended, i. e. including service type as another variable. Possible new admission policies also need to be proposed and analyzed. Since the performance of one service and the performance of the other services may not be independent in multimedia systems, analysis and design of admission control policies favoring handoff calls (or a particular service) will become more challenging. There are still many open questions on multimedia admission control. In Chapter 4, admission control policies in multimedia cellular systems are studied.

2. 3. 3 Modeling packet arrival process

In packet radio systems, both MS and BS transmit packets via wireless channels. Random multiple access is one of the access protocols used in the reverse link. Under this protocol, active MSs transmit their packets to a BS without coordination among themselves. The total traffic load is determined by the number of active MSs and the average number of packets generated by each MS per unit time. There are several models to describe the statistical characterization of the packet arrival processes. Since different services may generate packets in different manners, the models of packet sources can be different as well.

- Poisson modeling: In this model, the number of active MSs is assumed to be infinite. The probability of transmission per unit time per MS approaches zero [54]. Each user independently generates packets. The number of packets generated during certain period of time, T, follows Poisson distribution. The interarrival time between packets is negative exponentially distributed. When the number of active users in a cell is limited, say N, the binomial distribution can be used. In other words, the number of packets generated during certain period of time, T, follows a binomial distribution [37]. As N approaches infinite and each MS's contribution approaches 0, the binomial distribution approaches Poisson distribution. The advantages of this model are its analytical tractability and its simplicity in calculation. But it does not include the correlation among successive packets as observed within some services, such as voice service. It is suitable for modeling services without packet correlation but it is still used to model voice traffic by some authors. It is one of the widely used models in analyzing the performances of DS/CDMA cellular systems [20], [39], [61], [62]. It is a good start point to analyze system performance with this model. If correlation is considered, the following models can be used.
- Markov chain modeling: This model is proposed for the packet voice service [64]. Since a user talks and stops talking alternatively during a call, the appropriate model becomes a two-state discrete-time Markov chain. The transition probabilities between the ON state (the user is talking) to the OFF state (the user is silent) are listed in Table

2. 2 with the starting states in the left column. It is easy to show that the state probabilities of this Markov chain are: $p_{ON} = \beta/(\alpha+\beta)$ and $p_{OFF} = \alpha/(\alpha+\beta)$. While the user is in the OFF state, there is no packet being generated. Upon the user turns to the ON state, a number of packets will be generated and transmitted. Therefore, packet correlation is included in this model and the model is suitable for services with packet correlation. The model is more complex than the previous one. Moreover, different assumptions are made on the number of packets generated each time user turning to the ON state. In [65] and [66], binomial model is used; while in [31], the arrival is a fixed rate process. The two-state Markov chain needs to be extended to an N-state Markov chain to describe the case of N active voice users in a cell [64], [67]. The advantage of this model, of course, is the packet correlation being included. Its disadvantage is a higher calculation complexity, especially when N is large.

Table 2. 2: Transition probabilities of a two state Markov chain

	ON	OFF
ON	$1-\alpha$	α
OFF	β	$1-\beta$

In conclusion, both models are often used in packet cellular network analysis. When services with and without packet correlation are in one system, different models can be used together. The reason of choosing the binomial model in Chapter 3 is to reduce the computation complexity.

2. 3. 4 Markov chain

A stochastic process is a family of random variables, defined on a given probability space, indexed usually by the time t. If the state space of a stochastic process is discrete, the process is call a discrete-state process or simply a chain. Markov process is a special type of stochastic process whose future development is dependent only on its present state but not the history of its previous states. Markov processes can have a discrete state space or a discrete index space or both, which determines their type.

Among all types of Markov processes, we are interested in continuous-time Markov Chain, i. e. its index (time) space is continuous while its state space is discrete. It can be used to model a system having a number of discrete states. Assume that the system begins at $t = 0$. Let $X(t)$ be its state at time t, then it is a stochastic process:

$$X(t)\ ,\ \text{for}\ 0 \le t < \infty . \tag{2.12}$$

In such a chain, the transition from one state to the other can happen at any instant of time. Most important, the future evolution of a Markov chain depends only on its present state but not previous states. In other words, for any given time instants $t_0 < t_1 < t_2 < \ldots < t_n < t$, we have:

$$P(X(t) = x|_{X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \ldots X(t_0) = x_0}) = P(X(t) = x|_{X(t_n) = x_n}) . \tag{2.13}$$

where x_i is a state of the Markov process. Its state space is denoted as S with a limited or unlimited but countable number of states. Each and every individual state can be labeled by an integer. A continuous-time Markov chain is said to be irreducible if the chain can transit from any state to any other state. The state probability of a particular state i is defined as:

$$P_i(t) = P(X(t) = i) = P(i) . \tag{2.14}$$

Here we only consider the so called time-homogeneous Markov chain whose transition probabilities only depend on the time difference but not absolute time. The transition probability $P(t, x|y)$ is the probability that the system jumps from state y to state x, where t, from now on, denotes the time difference between x and y ($x \in S$ and $y \in S$). The behavior of such a Markov chain is completely determined upon the transition probabilities and the initial state probability are given. It is obvious that:

$$\sum_{x \in S} P(t, x|y) = 1\text{, and } \sum_{x \in S} P(x) = 1 . \tag{2.15}$$

where $P(0, x|y) = 0$ and $P(0, x|x) = 1$ $(x \neq y)$. The infinitesimal transition rates of the chain are defined as:

$$\gamma(x|y) = \frac{P(dt, x|y)}{dt}. \tag{2.16}$$

It is also found that the time of the system staying at a state is independent from that state and the staying time follows an exponential distribution law. A steady-state probability of state x is defined as:

$$\lim_{t \to \infty} P(t, x|y) = q(x), \text{ for } x \in S. \tag{2.17}$$

The above limit is always converges at $t \to \infty$ and does not depend on initial state y for irreducible Markov chain. To find the unknown $q(x)$, a set of linear functions can be set up, based on the Kolmogorov - Feller forward equations, as:

$$\sum_{y \neq x} \gamma(x|y) q(y) = \sum \gamma(y|x) q(x), \text{ for } x \in S \tag{2.18}$$

where $y \neq x$. The non-negative numbers $q(x)$ which always satisfy the system of linear equations (2.18) can be found. Obviously,

$$\sum_{x \in S} q(x) = 1, \tag{2.19}$$

always holds.

A special case of the continuous-time Markov chain is the so called birth-death process. In such a process, state transition only happens between the nearest neighbors. There exist non negative constant rates which are not functions of the time. The birth rate, λ_i, is the rate at which birth occurs in the state i such that the state of the chain is changed from state i to state $i+1$. Similarly, the death rate, μ_i, is the rate at which the state is transformed from i to $i-1$. It should be noticed that the above mentioned rates are not transition probabilities so they can be greater than 1. In any state, the occurrences of birth and the occurrences of death are independent. A birth-death process can be described by a set of linear equations like Eq. (2.18) as well as a state diagram. By solving those equations, all $q(x)$ can be obtained. More details about Markov chains and the birth-death processes can be found in literature [36], [68].

Due to the nature of the call arrival process in a cell, the behavior of the channel occupancy within a cell can be modeled by a birth-death process. The birth rate λ_i and death rate μ_i are considered to be stationary. The call arrival and departure rates are not changed with time. The birth-death processes are used in Chapter 4 to analyze admission control. In addition, non-birth-death processes are also used in Chapter 4. In these processes, the state transitions can happen even beyond nearest neighbors. They can be analyzed by a similar approach used on birth-death processes.

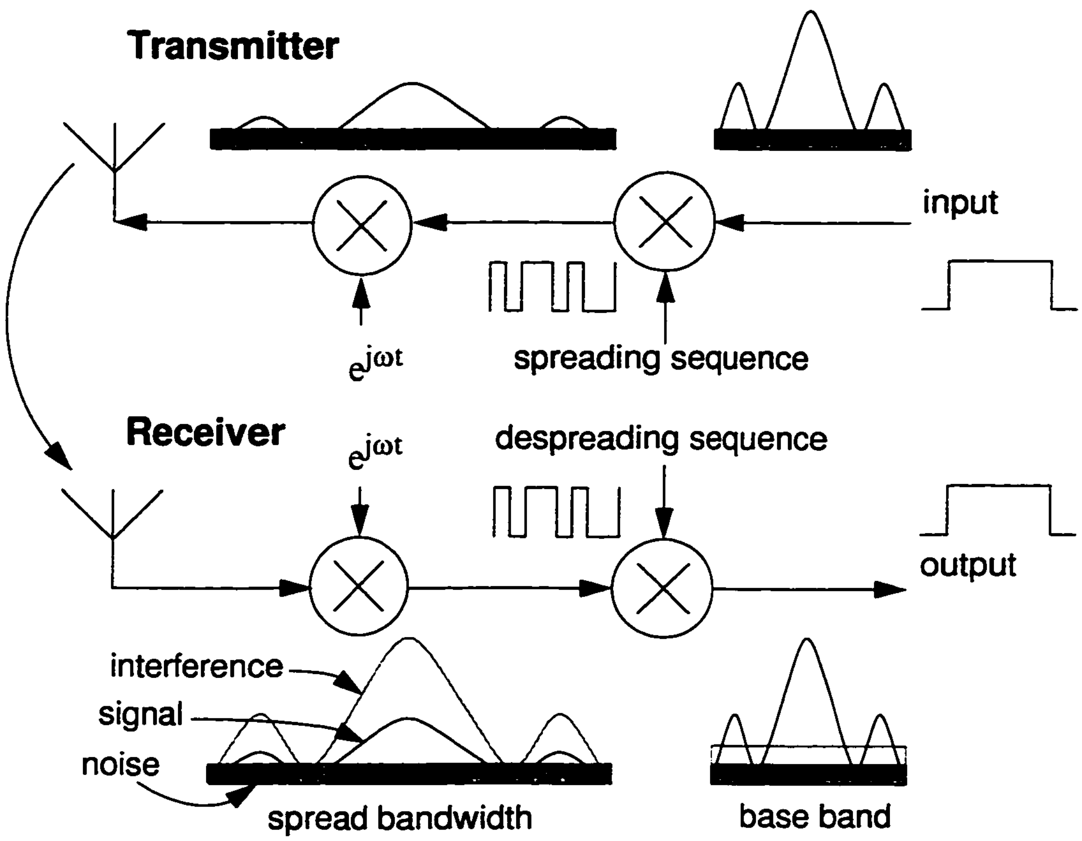


Fig. 2. 1 A simplified typical spread spectrum communication system. Interference seen by each user is from others signal.

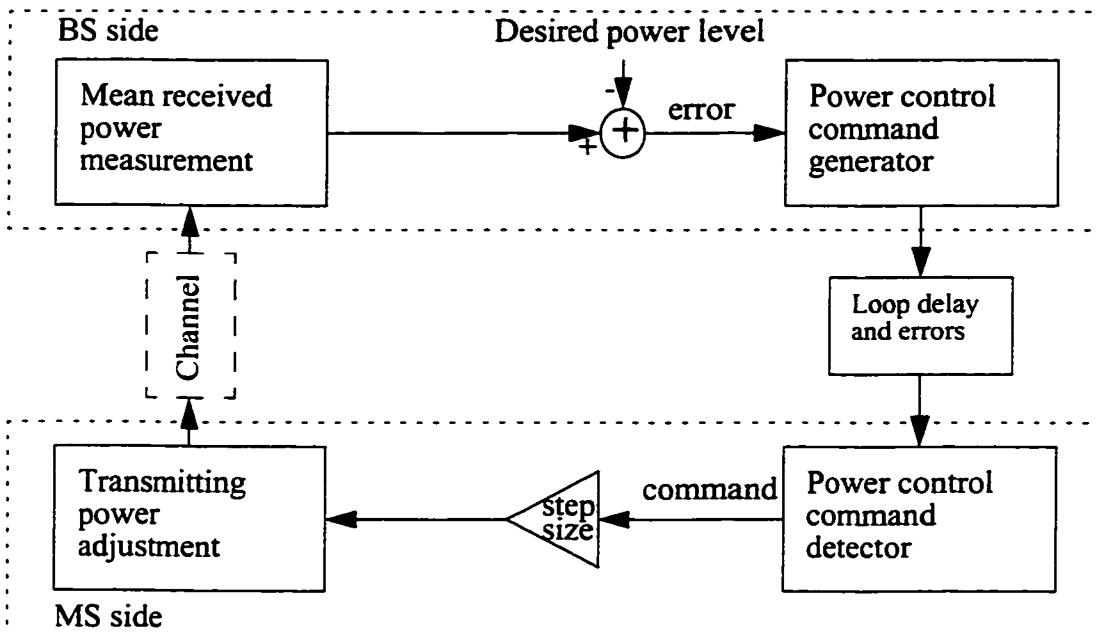


Fig. 2. 2 A typical closed loop power control model for CDMA cellular networks. Transmitting power level is updated every T seconds.

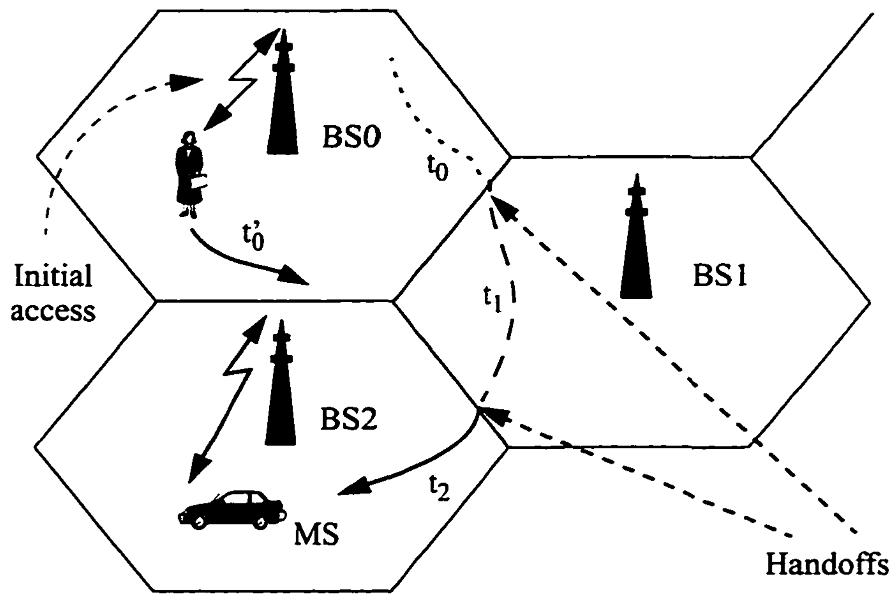


Fig. 2. 3 A typical wireless cellular system with users travelling among cells.

Chapter 3

Capacity Analysis of Multimedia CDMA Cellular systems

3. 1 Introduction

In DS/CDMA systems, power control mechanism keeps the system capacity on both directions as high as possible. Forward link power control is easier to accomplish due to the absence of the near-far effect [51], [52]. Reverse link power control must eliminate most of the near-far effect. By reverse link power control, MS adjusts its transmission power to a level that just keeps the QoS requirement of its service. Each MS is power controlled by a BS it is associated with. For multi-service CDMA systems, wider bandwidth allocation as well as more complicated power control and traffic management algorithms are expected [22], [27]. The requirement of efficient service integration introduces new problems in performance analysis and system design [23], [26]. It is necessary to obtain knowledge of such systems to find appropriate power control laws to improve their performance. In such systems, power control is still used to combat the near-far effect. Moreover, user's signal power received by a BS must follow certain power allocation rule (power control law and power allocation are used interchangeably in this chapter). Power allocation acts as a way of service integration in order to maximize the capacity and meet QoS of all services. It copes with the dissimilarity of services as discussed later. Power allocation and power control error are important factors affecting the capacity of a multi-service CDMA system.

Packet access is very desirable for systems carrying heterogeneous traffic [31]. It provides an efficient way of multimedia access. To control bit errors caused by poor wireless channels and cochannel interference, certain error control technology is used. To recover the corrupted bits in packets, either retransmission or error correction coding can be used. In this chapter, we focus on the modeling and evaluation of the reverse link performances of a DS/CDMA system with packet services.

3. 2 Analytical Model

3. 2. 1 Packet wireless DS/CDMA system model

We consider a slotted packet DS/CDMA cellular system supporting multi-service, i. e. type-1, type-2, ... , and type-Ω service. The time domain is divided into slots with length equal to the transmission time of a packet as shown in Fig. 3. 1. Packets of all services are of the same fixed length. They may overlap during transmission. The maximum waiting time of a packet, the duration between its generation and its transmission, is less than one slot. Each packet is spread over the entire system bandwidth W_s.

Since all MSs share the same bandwidth, the mutual interference limits the overall capacity. The total interference is composed of intercell and intracell interferences. For a single service system, it has been shown that when MSs are uniformly distributed within all hexagonal cells and the path loss is proportional to the 4th power of the MS-BS distance, the ratio of intercell interference to the intracell interference is about 33% [24]. For a multi-service system, it is also assumed that MSs of each service are uniformly distributed in all cells. Same pass lose rule is assumed as well. Therefore, the ratio of intercell interference to intracell interference of each service will be the same (i. e. 33%). Since the total intracell (intercell) interference is the summation of the intracell (intercell) interference of all services, the total intercell interference is also $f \approx 0.33$ times the total intracell interference.

To analyze the system, the i-*th* service, called type-i service, is chosen as a reference service to obtain the general results applied to all services. A parameter with a subscript i, is related to this service unless otherwise stated. We assume that the signal power of an MS (the MS is using type-i service) received by a BS, P_i, is log-normally distributed with pdf:

$$f_i(P_i) = \frac{1}{\sqrt{2\pi}\sigma_i P_i}\exp\left[\frac{(\ln P_i - m_i)^2}{2\sigma_i^2}\right], \text{ for } i = 1, 2, \ldots, \Omega \tag{3.1}$$

where σ_i and m_i represent the power control error and the logarithmic mean of P_i (the power level which is determined by power allocation), respectively. Obviously, σ_i, m_i and P_i are service dependent.

If v_i of u_i active MSs of the type-i service in a cell are transmitting in the same slot, the total signal power received by the BS in that slot is:

$$P_t = \sum_{i=1}^{\Omega}\sum_{j=0}^{v_i} P_{i,j}, \tag{3.2}$$

where $P_{i,j}$ is the power of the j-th MS of the type-i service received by the BS. The j-th MS, has a signal to interference ratio (SIR) at the BS as:

$$r_i = \frac{P_{i,j}}{P_t - P_{i,j} + I + n_o} = \frac{P_{i,j}}{P_t(1+f) - P_{i,j} + n_o}, \tag{3.3}$$

where $P_t - P_{i,j}$ is the total intracell interference, I is the total intercell interference and n_o is Gaussian noise. In the following analysis, n_o is neglected for simplicity since it is much smaller than the total interference. A special case of Eq. (3.2) is $v_i = 0$. It means that no type-i service MS is transmitting ($P_{i,0} \equiv 0$) in the slot. For a given u_i, v_i changes slot by slot, resulting in r_i becoming a random variable whose distribution is discussed later.

Each service has its own requirements on the BER and packet delay. To keep the BER under a certain value at the BS, a minimum SIR, $\Psi_i = (1+f)f_{sect}(E_b/N_0)_{thi}/PG_i$, must be met for the type-i service [21]. Here $(E_b/N_0)_{thi}$ is the bit energy to interference density ratio threshold. PG_i is the processing gain. Sectorization is reflected by a factor f_{sect} ($f_{sect} = 1$ means no sectorization and $f_{sect} = 1/3$ means three sectors per cell).

Since retransmission increases the packet delay, its suitability for a service is dependent on the delay requirement of the service. Some services, such as voice, are sensitive to packet delay but not BER. They have to be delivered on a real-time base. Bad packets are not retransmitted in order to meet the delay requirement. A service outage occurs whenever $E_{bi}/N_0 < (E_b/N_0)_{thi}$. It is required that the outage probability of a service should be lower than a threshold. Since a time slot is short, say $T = 20$ ms, the waiting delay introduced by slotted ALOHA is assumed to be acceptable. In microcell cellular network with high speed circuitry, additional delays due to transmission and processing are limited. Therefore we assume that all services without packet retransmission meet their delay requirements.

Other services, however, can tolerate relatively longer delays but require a quite low BER. For those services, if a packet is found to have an unacceptable BER due to $E_{bi}/N_0 < (E_b/N_0)_{thi}$, retransmission(s) is needed until an acknowledgment is received from the BS. It is assumed that the feedback channel carrying acknowledgments is error free. The performance metrics of these services are throughput and packet delay (not outage probability). Low BER requirements can be achieved by having packet retransmission and a higher $(E_b/N_0)_{thi}$. Services without retransmission have priority over the ones with retransmission to control the number of lost unrecoverable packets. The amount of mixed traffic has to be controlled to keep the outage probability and the packet delay acceptable.

3. 2. 2 Packet generation model

In each time slot, for type-i service, there are u_i MSs generating v_i packets with a probability of $p_i(u_i, v_i)$. If the service has no retransmission, its packet generation can be modeled by a binomial distribution:

$$p_i(u_i, v_i) = \binom{u_i}{v_i} \alpha_i^{v_i} (1-\alpha_i)^{u_i - v_i}, \tag{3.4}$$

where α_i is the activity factor of the type-i service. If the service uses a retransmission protocol, we assume that its offered traffic G_i, including new and retransmitted packet, is the average number of packets per slot generated by u_i MSs. Then the packet generation model can be written as:

$$p_i(u_i, v_i) = \binom{u_i}{v_i}\left(\frac{G_i}{u_i}\right)^{v_i}\left[1-\frac{G_i}{u_i}\right]^{u_i-v_i}. \qquad (3.5)$$

The number of packets from all services in a slot is denoted as $\boldsymbol{v} = [v_1, v_2, \ldots, v_\Omega]^T$. Assuming each MS generates packets independently, the probability that MSs $\boldsymbol{u} = [u_1, u_2, \ldots, u_\Omega]^T$ generate $\boldsymbol{v}$ packets in a time slot is:

$$p(\boldsymbol{u}, \boldsymbol{v}) = \prod_{i=1}^{\Omega} p_i(u_i, v_i) \qquad (3.6)$$

If there are too many packets in one slot, $(E_b/N_0)_i$ may be lower than $(E_b/N_0)_{thi}$. If that happens, we assume that all packets with $(E_b/N_0)_i < (E_b/N_0)_{thi}$ are destroyed, which means those packets either being lost or waiting for retransmission.

3. 3 System Capacity

3. 3. 1 Capacity of mixed traffic

Ongoing calls interfere with each other. In order to calculate the system capacity under given outage probability and delay requirement, we need to know the distribution of the interference from all other users on one user. Then the distribution of r_i, which is the same for all users of the type-i service, can be determined. By using the method proposed in [21], [23], we first derive the mean and variance of the total intra-service, intra-cell interference to a user from all other users of the type-i service. This interference distribution can be approximated by another log-normal random variable with variance $f_{\sigma i}$ and mean f_{mi}, as proved by Fenton [23]. $f_{\sigma i}$ and f_{mi} are function of v_i, m_i and σ_i as:

$$f_{\sigma i}(v_i, \sigma_i) = \ln(e^{\sigma_i^2} + v_i - 1) - \ln v_i \qquad v_i > 0, \qquad (3.7)$$

$$f_{mi}(v_i, m_i, \sigma_i) = \frac{1}{2}\left[3\ln v_i + 2m_i + \sigma_i^2 - \ln(e^{\sigma_i^2} + v_i - 1)\right] \qquad v_i > 0, \qquad (3.8)$$

When there are $\sum v_i$ packets in one slot, to a particular type-i packet, $v_i - 1$ other type-i packets and all the rest of packets are interference. Using the approach in [30], we obtain the pdf of the total interference by the following steps.

- Step 1: combine the interference of type-1 service with that of type-2 service and calculate a summation (partial total interference) with its mean and variance. Appendix B gives the algorithm of combining.
- Step 2: combine the interference of type-3 with that of the summation to obtain a further summation.
- Step 3: repeat Step 2, i. e. add a service at a time until all Ω services are included in the interference calculation.

The total interference is also log-normal according to Appendix B. Assuming that the total interference and the user signal are independent, the pdf of r_i of a type-i packet is:

$$f_{SIR_i}(r_i) = \frac{1}{\sqrt{2\pi(\sigma_{Ii}^2 + \sigma_i^2)}\, r_i} \exp\left[-\frac{(\ln r_i - m_{Ii} + m_i)^2}{2(\sigma_{Ii}^2 + \sigma_i^2)}\right], \tag{3.9}$$

where m_{Ii} and σ_{Ii}^2 are mean and variance of the total interference to the type-i packet, respectively. They are complicated functions of v, m_i and σ_i, and can be obtained numerically. Based on the pdf, we have the packet failure probability of the type-i service packets as:

$$p_{fi} = \int_0^{\Psi_i} f_{SIR_i}(r_i)\, dr_i \text{ for } i = 1, 2, \ldots, \Omega. \tag{3.10}$$

After some mathematical manipulations, the closed form of the packet failure probability becomes:

$$p_{fi} = \frac{1}{2} + \frac{1}{2} erf\left((\ln \Psi_i - m_{Ii} + m_i) / \sqrt{2(\sigma_{Ii}^2 + \sigma_i^2)} \right), \tag{3.11}$$

where $erf()$ is the error function defined as:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \tag{3.12}$$

It is clear that the packet failure probability is a function of power allocation, power control error and the number of packets of each service in a slot. Now we can calculate the performance of each service based on packet failure probability and error control method. If the type-i service has no packet retransmission, its performance measurement is the average packet outage probability. When there are $\boldsymbol{u}$ active users in a cell, the probability is given by:

$$P_{oi} = \sum_{v_1 = 0}^{u_1} \cdots \sum_{v_\Omega = 0}^{u_\Omega} p_{fi} p(\boldsymbol{u}, \boldsymbol{v}) \,. \tag{3.13}$$

However, if the type-i service allows packet retransmissions, its performance metrics are packet throughput and average packet delay. For u_i type-i service MSs, their offered traffic is:

$$G_i = \alpha_i u_i \,, \tag{3.14}$$

where α_i is activity factor of the service. When there are $\sum u_i$ MSs in a cell, the whole throughput of type-i service is given by:

$$S_i = \sum_{i_i = 1}^{u_i} i_i \left[1 - \sum_{i_1 = 1}^{u_1} \cdots \sum_{i_\Omega = 0}^{u_\Omega} p_{fi_i} \prod_{k = 1, k \neq i_i}^{\Omega} p_k(u_k, i_k) \right] p_{i_i}(u_{i_i}, i_{i_i}) = \sum_{i_i = 1}^{u_i} i_i [1 - \overline{p_{fi_i}}] p_{i_i}(u_{i_i}, i_{i_i}) \,. \tag{3.15}$$

where $\overline{p_{fi}}$ is the average probability that a packet is unsuccessfully transmitted. The summation to obtain $\overline{p_{fi}}$ in Eq. (3.15) does not include a summation over i_i. With G_i and S_i obtained, the average packet delay, defined as the time between the generation and the correct reception of a packet, and normalized to a slot duration, is [21]:

$$D_i = 1.5 + \left(\frac{G_i}{S_i} - 1 \right) (\lfloor E(\tau_i) + 1 \rfloor + 1) \,, \tag{3.16}$$

where $E(\tau_i)$ is the mean of the retransmission delay, which depends on the cell size and the processing speed of the type-i service.

We then define the capacity as a set of vectors in the form of $[n_1, n_2, \ldots, n_\Omega]$. It is a set of user combinations such that the QoS of all services are met and adding one more MS of any service will result in not all QoS being met. It is no longer a number as that in a single traffic system. The capacity is:

$$C = \{max(u_1, u_2, \ldots, u_\Omega) | \boldsymbol{P_o} \le \boldsymbol{P_{th}}, \boldsymbol{D} \le \boldsymbol{D_{th}}\} = \{ [n_1, n_2, \ldots, n_\Omega] \} , \qquad (3.17)$$

where $\boldsymbol{P_o} = [P_{o1}, P_{o2}, \ldots, P_{o\Omega}]^T$ is the outage vector with the component P_{oi} as the outage threshold of the type-i service. $\boldsymbol{D} = [D_1, D_2, \ldots, D_\Omega]^T$ is the average packet delay vector with a vector $\boldsymbol{D_{th}} = [D_{th1}, D_{th2}, \ldots, D_{th\Omega}]^T$ as its threshold. Notice that the symbol $\le$ in Eq. (3.17) means that $P_{oi} \le P_{thi}$ and $D_i \le D_{thi}$ hold for all i. The maximum average number of type-i packets per slot that n_i MSs generated is $G_{mi} = \alpha_i n_i$.

$max(u_1, u_2, \ldots, u_\Omega)$ in Eq. (3.17) means special u_i combinations, denoted as $[n_1, n_2, \ldots, n_\Omega]$. For a $[n_1, n_2, \ldots, n_\Omega]$, if we increase any n_i by 1, we must reduce at least one n_j $(j \ne i)$ to hold $\boldsymbol{P_o} \le \boldsymbol{P_{th}}, \boldsymbol{D} \le \boldsymbol{D_{th}}$. For other u_i combinations, we can find at least one u_i such that $\boldsymbol{P_o} \le \boldsymbol{P_{th}}, \boldsymbol{D} \le \boldsymbol{D_{th}}$ still hold even one more type-i service MS becomes active. The total number of $[n_1, n_2, \ldots, n_\Omega]$ can be high. All $[n_1, n_2, \ldots, n_\Omega]$ are elements of C.

The higher the $\boldsymbol{P_{th}}$ and/or the longer the $\boldsymbol{D_{th}}$, the higher the C. If $G_i > G_{mi}$ (i.e., $u_i > n_i$), it will result in that $P_{oi} > P_{thi}$ and/or $D_i > D_{thi}$. The rest of the services may be affected by the type-i service and their performance may also become unsatisfactory due to the type-i service interference increase. It should be noted that the value of n_i is highly dependent on α_i when we calculate C from Eq. (3.4), (3.5), (3.13), (3.16) and (3.17).

Soft capacity can be achieved if we temporarily tolerate $\boldsymbol{P_o} > \boldsymbol{P_{th}}$ and/or $\boldsymbol{D} > \boldsymbol{D_{th}}$ for certain service(s). As we loose the $\boldsymbol{P_{th}}$ and/or $\boldsymbol{D_{th}}$ requirements, the soft capacity $C_{soft} > C$.

3. 3. 2 Signal power allocation and system capacity

It is assumed that a strength-based closed loop power control mechanism tries to make all type-i service MSs having the same m_i at a BS. For convenience, we assume that the service requiring the lowest power allocation has a power level of unit (0 dB). The service is referred to as the type-1 service (i. e. $m_1 \equiv 0$ dB). All other services have higher power levels. Only the relative power level differences are of interest.

Power allocation affects system performance. Since the services use different error control schemes, the impact of power allocation on their performances might be different as well. Assume the type-i service has packet retransmission, for example. If m_i is too high, we have a very low $\overline{P_{fi}}$ and a very short D_i, but its throughput S_i might be low due to a small amount of G_{mi} allowed (fewer type-i calls can be supported in order to protect unrecoverable packets and to keep the delay of recoverable packets of other services acceptable). Therefore the capacity is reduced. On the contrary, if m_i is too low, we can have a large G_{mi} (more type-i calls can be admitted thus increasing the capacity). However, a higher $\overline{P_{fi}}$ limits S_i, resulting in a longer average packet delay D_i of its own. Even $D_i > D_{thi}$ might happen. Lots of low power packets are lost turning signal into harmful interference. To determine the proper power level of a service, QoS of all services must be taken into account since the services share the bandwidth and interfere with each other.

Therefore, when $\boldsymbol{D}_{th}$ and $\boldsymbol{P}_{th}$ are given, we can find an optimum power allocation plan which just guarantees all the QoSs and maximizes the capacity in terms of accommodating the largest number of calls simultaneously. For the type-i service, we further denote:

$$C_i = \{max(u_1, u_2, \ldots, u_\Omega) \,|\, P_{oi} \le P_{thi}, D_i \le D_{thi}\} = \{ [n_{1i}, n_{2i}, \ldots, n_{\Omega i}] \} , \qquad (3.18)$$

as the capacity that only QoS of the type-i service is guaranteed to be met and just above its threshold (QoS of the other services may or may not be met). Notice that when C_i is reached, some service user's signal quality may still below their corresponding threshold. C_i will change if we adjust power allocation and/or change service QoS threshold. The notation of Eq. (3.18) is similar to that of Eq. (3.17). The system capacity is related to C_i as:

$$C = min(C_1, C_2, \ldots, C_\Omega) . \qquad (3.19)$$

Here min means that only those $[n_{1i}, n_{2i}, \ldots, n_{\Omega i}]$ combinations, which satisfy Eq. (3.18) for *all* services, are chosen to be the elements of C. If the power allocation used results in that one (or more) particular C_i are far apart from the rest, the system resources is wasted. This power allocation plan should be avoided by further power level adjustment.

Power control error is another factor affecting the overall capacity and packet delay since it affects the interference statistics. In reality, there is always power control errors. Service with a small α_i is bursty in nature. Burst makes accurate power control more difficult, since the channel condition changes between consecutive packets bursts which is difficult to track. Therefore, different services may show different power control errors. Moreover, the sensitivity of each service to the power control error may be different too. The impact of power control error on the system capacity is shown later.

We finally consider a special case. There are two services carried by the system. One is voice service without packet retransmission. The other is data service with packet retransmission. The data service has a higher source data rate, say twice as that of voice service. There are two schemes to transmit the high-rate service. In Scheme 1, m_2 is increased and PG_2 is reduced by half. The packet length is kept as one slot. In scheme 2, each packet is divided into two and transmitted in two consecutive time slots with a lower m_2 after the packet is generated. PG_2 is unchanged. In scheme 2, the probability of n_2 data users generating k_2 (where $0 \le k_2 \le 2n_2$) packets in a slot is [30]:

$$p'_2(n_2, k_2) = \sum_{k_1 = max(0, k_2 - n_2)}^{min(k_2, n_2)} \binom{n_2}{k_2 - k_1}\binom{n_2}{k_1}\left(\frac{G_2}{n_2}\right)^{k_2}\left(1 - \frac{G_2}{n_2}\right)^{2n_2 - k_2} \tag{3.20}$$

$p'_2(n_2, k_2)$ is used to calculate P_{out}. The throughput in Scheme 1 is given by Eq. (3.15). The throughput in Scheme 2 becomes:

$$S_2 = \frac{1}{2}\sum_{j=1}^{2n_2} j[1 - \overline{p_{f2}}(j)]p'_2(j) \tag{3.21}$$

Since half of the information is delayed by one slot, the average packet delay becomes:

$$D'_2 = D_2 + 0.5. \tag{3.22}$$

We will compare the delay performances of the schemes in the following section.

3. 4 Numerical Results and Discussions

Systems carrying two services are considered, i. e. $\Omega = 2$. In the following discussions two cases are provided. In the *case 1*, voice service, (the type-1 service without retransmission) plus data service (the type-2 service with retransmission) are carried by the system. In *case 2*, two data services both allowing retransmission share the entire system bandwidth.

In the numerical examples we assume the following service parameters for both cases. For data service(s), source data rate of 9.6 kbps ($PG_2 = 256$) is considered. We require $(E_b/N_0)_{th2} = 12$ dB which enables $BER \approx 4 \times 10^{-6}$ [65]. Voice service is transmitting at 9.6 kbps with $PG_1 = 256$ and $(E_b/N_0)_{th1} = 7$ dB to keep $BER \leq 10^{-3}$ [20]. The total system bandwidth is 2.5 MHz.

3. 4. 1 Case 1

In the first case, it is assumed that $\alpha_1 = 3/8$ and $\alpha_2 = 0.1$. We chose that the power level of voice service is unity ($m_1 = 0$ dB). Since data users require a lower BER, their power is higher than unity at the BS. We seek to see the mutual influence between services.

First, the impact of the difference between m_2 and m_1 on the capacity is examined. Fig. 3. 2 (a) shows that when $m_2 - m_1$ increases, C is reduced because the interference from data service to voice service increases. However, a higher m_2 can improve the performance of data service due to a relatively lowered voice service interference (see Fig. 3. 3).

The difference between m_2 and m_1 also affects S_2 and D_2. From Fig. 3. 3, we see that a higher m_2 results in higher S_2 and shorter D_2. The cost of S_2 improvement is a lower capacity clearly shown in Fig. 3. 2. When m_2 is low and n_1 is decreasing, D_2 will reduce and S_2 will increase slightly faster with G_2 because of lower voice interference. Therefore, if we only need to allocate a rather low power to type-2 service to keep its QoS, its delay will be significantly affected by the population of type-1 service users. The bottom line is to allocate enough power to type-2 service such that its QoS is still met under the largest possible type-1 population. Of course, the selection of m_2 is based on a given D_{th}. For example, if D_{th2} = 120 ms, then we let m_2 = 2 dB. There are two important observations. First, the decrease of D_2 becomes slower as m_2 becomes higher. Second, type-2 service with a high enough m_2 has a shorter D_2 which is less sensitive to type-1 service population variation.

Our results in Fig. 3. 2 (b) show the impact of power control error on C. First, when σ_2 and σ_1 increase from 1 to 3 dB, C is greatly reduced. For a higher σ_1, more voice packets are weaker (Of course, more packets are stronger as well. However, only the weaker ones limit the performance). This results in a very small capacity, which corresponds to a very small mutual interference, in order to protect those weak voice packets to satisfy the $P_{o1} < P_{th1}$ requirement. Since corrupted voice packets are not going to be retransmitted, we only can protect them by limiting mutual interference. Second, for the service allowing retransmission, a higher power control error σ_2 also reduces C. However, the capacity sensitivity on σ_2 and σ_1 is different. Capacity under σ_2 = 1 dB and σ_1 = 3 dB is much smaller than under σ_2 = 3 dB and σ_1 = 1 dB. Finally, the capacity is also more sensitive to the σ of a service having higher amount of traffic. In Fig. 3. 2 (b) when σ_2 = 3 dB (σ_1 = 1 dB), the capacity loss becomes smaller and smaller as n_2 decreases. In general, tight power control is necessary in Case 1. On contrary, in systems in which only data service is carried, higher σ can result in higher throughput under heavy traffic due to capture [29]. For a well power controlled service, packets of similar power level are more likely to be destroyed by collisions rather than to be captured by a BS. For a poor power controlled service, high power packet can get through.

Power control error also affects S_2 and D_2. When σ_1 = 1 dB, as shown in Fig. 3. 4, S_2 and D_2 are affected by σ_2. First, when m_2 is low, increased σ_2 helps increase S_2 and reduce D_2. The reason is that more high power level packets are generated due to power control error and they can be captured by a BS. Second, when m_2 is high, high σ_2 reduces S_2 and C. However, D_2 is nearly unchanged. Poorly power controlled high power level type-2 service generates more fluctuating interference such that we have to limit its population to protect type-1 service. So it is important to have a better power control over high power level data service. In addition, for σ_1 = 3 dB, the capacity loss makes system performance unacceptable. Minimizing σ_1 is very important. In summary, proper power allocation plus tight power control improve overall system performance in terms of capacity, delay and throughput.

Fig. 3. 5 (a) shows the relationship between the maximum value of S_2: S_{m2}, and m_2 for given n_2 and n_1. Here we always keep $P_{o1} \le P_{th1}$. To satisfy this condition, there exists an maximum achievable S_2 for a given MS population and power allocation plan. We vary α_2 of the data service until a maximum S_2, denoted as S_{m2}, is obtained under a given m_2. It is found that S_{m2} begins to drop from its maximum as m_2 reaches certain point (about 3 dB in Fig. 3. 5 (a)). Therefore, there is a limit of improving the data service performances by means of raising m_2. Moreover, S_{m2} decreases as n_1 increases, i. e. voice service takes more and more capacity. Fig. 3. 5 (b) indicates that delay reduction becomes slower as m_2 increases, so the data service's gain in delay reduction can not justify the system capacity loss when m_2 is too high.

Fig. 3. 6 shows that P_{o1} is quite sensitive to the power allocation. P_{o1} increases only slightly faster as u_2 increases if m_2 is higher. Under light traffic, higher m_2 reduces packet delay as in Fig. 3. 7. However, when traffic is heavy, higher m_2 increases packet delay under a given throughput due to its intra-service interference. At the cost of higher P_{o1}, the maximum packet throughput is also increased as m_2 becomes higher in Fig. 3. 7.

Now we look at the case of high rate data service, i. e. the source date rate of data service is 19.2 kbps. Two mentioned schemes (scheme 1: $PG_1 = 128$; scheme 2: $PG_2 = 256$) are analyzed and compared. By choosing appropriate m_2, both schemes have nearly the same C as shown in Fig. 3. 8 (a). Then their achievable delay performances are compared in Fig. 3. 8 (b). When m_2 is low, D_2, in scheme 2, is generally shorter than that in scheme 1. D_2 is also less sensitive to the variation of n_2 which reflects the amount of voice traffic. Since in scheme 2 PG_2 is high, data packets are more robust towards the interference of other services. Therefore, scheme 2 is more suitable for systems with longer D_{th2}, and provides another mean of service integration. The advantage of scheme 1 is its easier implementation. Because the improvement on D will shrink as m_2 increases, scheme 1 can be used in systems requiring a rather short D_{th2}.

3. 4. 2 Case 2

For the case of two data services, in order for us to focus on the impact of power allocation and power control error on throughput, delay and capacity performances, we assume that $(E_b/N_0)_{th1} = (E_b/N_0)_{th2} = 12$ dB and $\alpha_1 = \alpha_2 = 0.3$. Performance in terms of capacity, throughput and delay under different power allocation and power control error are examined. The power level of type-1 service is unit (0 dB). However, D_{th1} and D_{th2} may be different.

First, throughput and packet delay are shown in Fig. 3. 9 (a) and (b) for both services (where $\sigma_1 = \sigma_2 = 1$ dB and no D_{thi} is considered). For a given u_1, S_1 gradually decrease as u_2 increases. At the same time, S_2 increases until reaching its top and begins to decline since intra-service interference of type-2 service causes more and more retransmissions. Therefore, admission control should be implemented to prevent system throughput from declining. We also notice that $S_2 > S_1$ under heavy traffic load due to $m_2 > m_1$. Service 1 is relatively blocked if the traffic is too heavy. From Fig. 3. 9 (b), we notice that the average packet delay difference between the two services in log scale, under a given $m_2 - m_1$, is not sensitive to the user number variation. The average packet delay increases in a similar fashion as the number of active users increases. It is also find that loosing power control benefits the throughput of both services under heavy traffic condition due to the capture effect.

The impact of power allocation (m_1 and m_2) and power control error (σ_1 and σ_2) on both throughput ratio and delay ratios of the two services (where $\boldsymbol{u} = [25, 25]^T$) are shown in Fig. 3. 10 (a) and (b). These ratios reflect the performance dissimilarity between services and how m_i and σ_i affect this dissimilarity. Under good power control ($\sigma_1 = \sigma_2 = 1$ dB), as m_2 increases, the average packet delay of the type-1 service increases very fast. However, when $\sigma_1 = 3$ dB, their performance dissimilarity on throughput and delay shrinks, i. e. both ratio curves are closer to the horizontal line of 1. Moreover, the case of $\sigma_1 = 3$ and $\sigma_2 = 1$ shows an interesting outcome. D_2 can be longer than D_1 even $m_2 > m_1$. It shows that to achieve a shorter D_2, a higher m_2 may not be an absolute guarantee. Another important observation is that, for given m_1 and m_2, variation in σ_1

causes greater change in both D_2/D_1 and S_2/S_1 than variation in σ_2 does. However, the trends are quite different (i. e. D_2/D_1 is more sensitive to σ_1 at smaller m_2, while S_2/S_1 is more sensitive to σ_1 at higher m_2). We can conclude that the system performance has different sensitivities towards σ_i and σ_i has a great impact on the degree of service performance dissimilarity. If necessary, we must select a power allocation which can compensate for the impact.

Finally, we look into the system capacity. When the $\boldsymbol{D}_{th}$ are known, the capacity of mixed traffic can be calculated. Fig. 3. 11 shows how the delay threshold affects the system capacity. It is not surprising that a tighter $\boldsymbol{D}_{th}$ reduces the system capacity C. In the case of $D_{th2} = 2.5$, the system capacity is fully determined by the type-2 service's constrains. For type-1 service, even when the capacity is increased to the solid line in the figure, its delay requirement is still satisfactory (but not type-2 service's). This means the waste of the system resources in terms of reduced capacity and unnecessarily good type-1 service quality. Same observation is found in the case of $D_{th2} = 6.0$. This time, type-1 service becomes the bottle neck. Obviously, it is desirable to have packet delays of both services approaches their thresholds nearly at the same time.

From the example in Fig. 3. 12, it is found that an appropriate power allocation exists and can solve above problem by means of minimizing $|C_2 - C_1|$. Usually, $C_2 \neq C_1$ and therefore, $C \neq C_1$ or $C \neq C_2$. It results in, for some combinations of (n_1, n_2), the QoS of one service is still well below its threshold and more users of that service can be accommodated, while the QoS of the other service is already above threshold. By increasing the power level of the service not going to meet its QoS, its quality is improved until meeting its QoS. In this figure, $m_2 = 1$ dB outperforms other two cases since $|C_2 - C_1|$ is smaller over different n_2. Therefore, to adjust the power level of the services could be helpful to maximize the system capacity. A specific packet delay requirement $\boldsymbol{D} < \boldsymbol{D}_{th}$ determines a power allocation plan and limits the number of active users in a cell.

3. 4. 3 Power control law

From Case 1 and 2, we can provide the following discussions. The power control law describes the relationship between system parameters and a power allocation plan. An optimum law maximizes the system capacity C described in Eq. (3.17). Notice that C is a very complicated function of power control law m_i, power control error, the QoS of services, as well as the traffic load of the services. Unfortunately, there is no explicit function found to show the relationship among m_i and the system parameters. However, We can seek the power control law of a high quality service based on its QoS under simplified cases.

For the two services in Case 1, the problem becomes finding m_2 for given quality parameter $\boldsymbol{D}_{th}$. From Fig. 3. 5 (b), it is found that the product of the value of delay and the value of m_2 is nearly a constant for $m_2 > 2$ dB, which suggests a simplified power control law: allocate power to type-2 service reversely proportional to its delay requirement. As data service requires very low BER, it should be allocated a rather higher power than voice to reduce packet retransmission. Thus we can apply the above law.

Furthermore, we discuss the power control law in Case 2. To simplify the problem, we normalize D_2 with D_1 to get the D_2/D_1 ratio and draw its relationship with m_2 in Fig. 3. 10 (a). This non-linear relationship is power control law. In other words, given $\boldsymbol{D}_{th}$, we can calculate D_{th2}/D_{th1} and find a corresponding m_2 in the figure. Since the absolute delay values are also concerned, the number of active users should be controlled to fulfill specific absolute delay value requirements of each service. In Fig. 3. 9 (a) we notice that D_2 and D_1 change approximately at the same rate, in a log scale, as the user population changes. That means D_2/D_1 is insensitive to user population variation. It is also noticed that when $\sigma_2 = 3$ dB, D_2/D_1 and m_2 nearly follow a linear relationship. We can still allocate power to type-2 service reversely proportional to the ratio of their delay requirements.

Table 3. 1: Outage probabilities of the two services[a]

m_2, Ψ_2 (dB)	1, 8	2, 9	3, 10	4, 11	5, 12
p_{o1}	7.7×10^{-4}	2.5×10^{-3}	7.3×10^{-3}	1.8×10^{-2}	3.9×10^{-2}
p_{o2}	7.2×10^{-4}	2.2×10^{-3}	6.2×10^{-3}	1.5×10^{-2}	3.1×10^{-2}

a. assuming $u_1 = u_2 = 40$, $\alpha_1 = \alpha_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$ dB, $m_1 = 0$ dB, and $\Psi_1 = 7$ dB.

Another interesting case is that two services in a system both require very short packet delay. They do not use packet retransmission. The outage probability is their QoS metric. We allocate a m_2 such that $m_2 - m_1 = \Psi_2 - \Psi_1$, i. e. the power allocated to a service is proportional to its quality requirement Ψ_i. Their outage probabilities are listed in Table 3. 1. The differences between these outage probabilities are quite small. This fact confirms the power control law proposed in [35]: to allocate power proportional to service quality requirement makes the outage probability of all services approximately the same.

For other cases, a search for power control law may be unavoidable. We start with analytical results of power allocation and then gradually adjust power of services which have either over good qualities or unacceptable qualities. To find an optimum power control law for more than two services, the search effort is dramatically increased. We should reduce the power control error as much as we can to increases capacity. The outage probabilities and/or packet delays under different power allocation are measured to see if the QoS of every service is met. Power adjustment is made based on those measures in order to gradually approach an optimum power allocation.

3. 5 Conclusions

In this chapter, we have discussed the reverse link capacity of cellular DS/CDMA systems with multi-service traffic. The services can either be with or without packet retransmission as an error control method. We first developed a model for analyzing the capacity, packet throughput and delay. Then affecting factors including power levels, power control errors, and delay requirements were discussed. Two numerical cases are given.

Our results on Case 1 show that increasing data service power level improves its performance but reduces the number of ongoing voice calls. Throughput and delay requirements determine the power level of data service. Capacity is more sensitive to the power control error of voice service. By appropriate power allocation, we can maximize capacity while still meeting their quality requirements. In Case 2, for given delay thresholds, there also exists an optimum power allocation maximizing the system capacity.

Therefore, we can make the following conclusions.

- The optimum power allocation plan is a compromise among dissimilar services. Increasing one service's power is always at the cost of the others. The relationship between power allocation plan and the QoS is complex, i. e. non-linear. Under some cases, simple power control law is found, in which power allocation of a high quality service is reversely propositional to its delay threshold.
- System performances are sensitive to power control accuracy, but the sensitivity is service and traffic amount of that service dependent. Power control overhead, such as the feedback channel bandwidth, given to a service may be selected based on how great the impact of its power control error on the system performance is.

  Power control error is a negative factor. Service with retransmission can use power control error to gain capture capability. But other services might be affected. As we do not allow cells overloading, we unlikely take advantage of capture capability.

  Since power control error has (sometimes great) impact on system performance, we should take this impact into account when choosing a power allocation plan.
- Power allocation plan should make all service qualities being only marginally kept above required thresholds when the system capacity is reached. Over-good service quality wastes system resources. Power allocation should be adjusted if this happens.

Power level assignment and power control are effective ways to integrate multi-service into one DS/CDMA cellular system. Beyond this, other control issues in multi-service scenario, such as admission control, are also important and will be addressed in our later work. Since power allocation and power control error can change the capacity of a system, the blocking probability of calls will also be affected as shown in the next chapter.

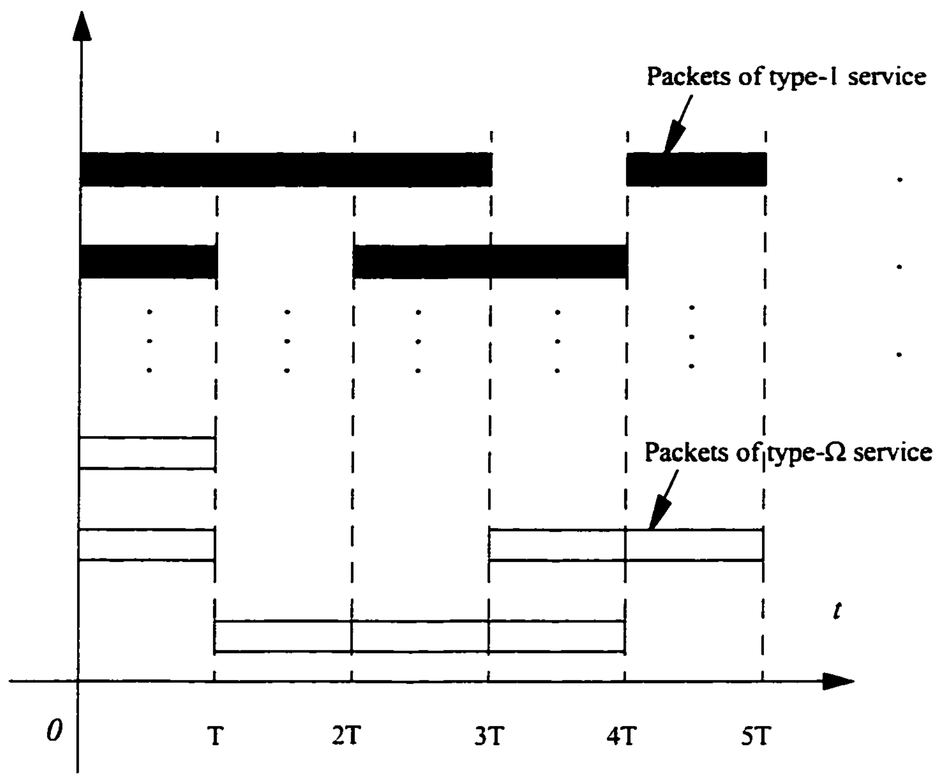


Fig. 3. 1 A slotted packet DS/CDMA system with voice and data traffic.

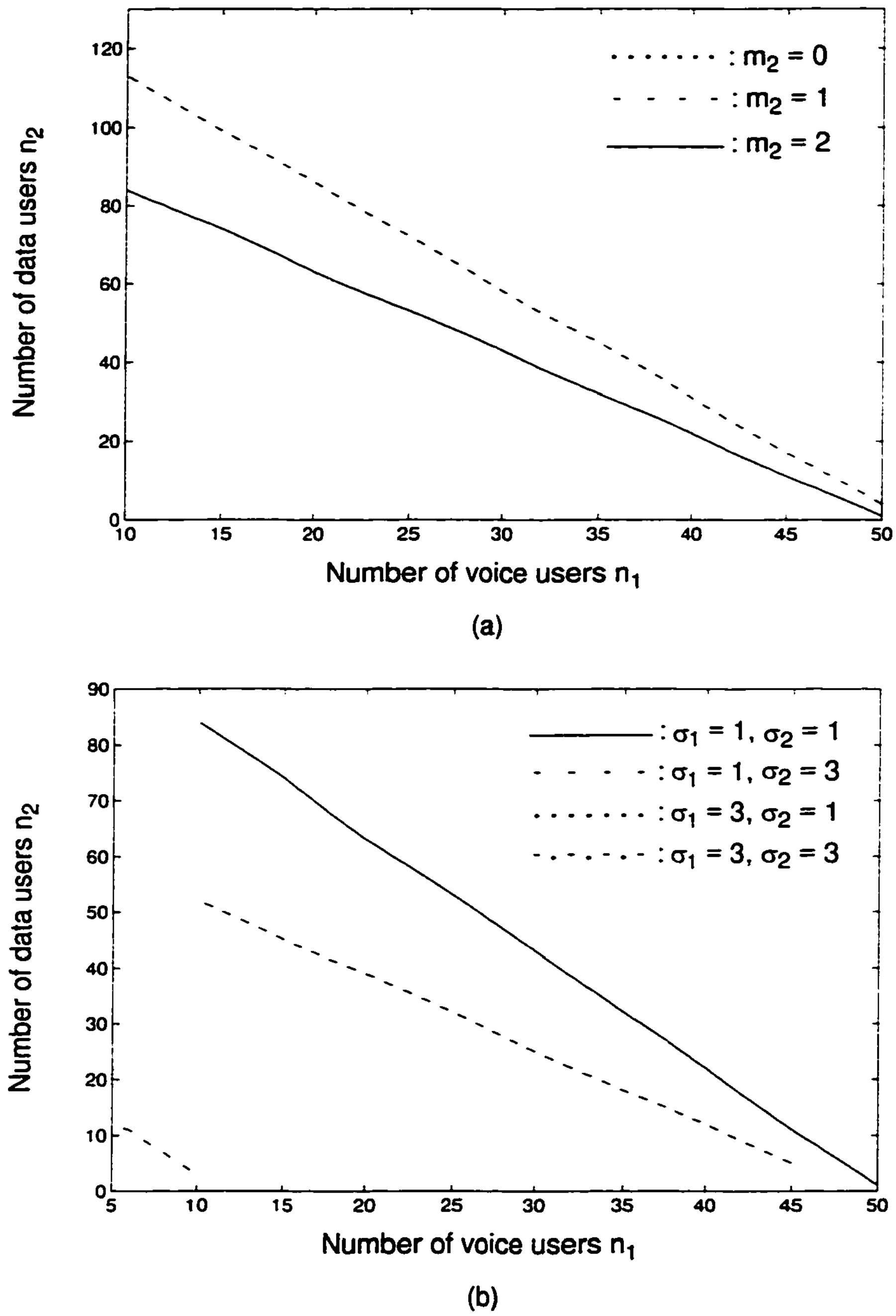


Fig. 3. 2 (a) Capacity of mixed traffic. Power level difference between data and voice service affects the overall capacity. $m_1 = 0$ dB. $\sigma_1 = 1$ dB, $\sigma_2 = 1$ dB. (b) Power control errors and their impact on the overall capacity. $m_1 = 0$ dB, $m_2 = 2$ dB.

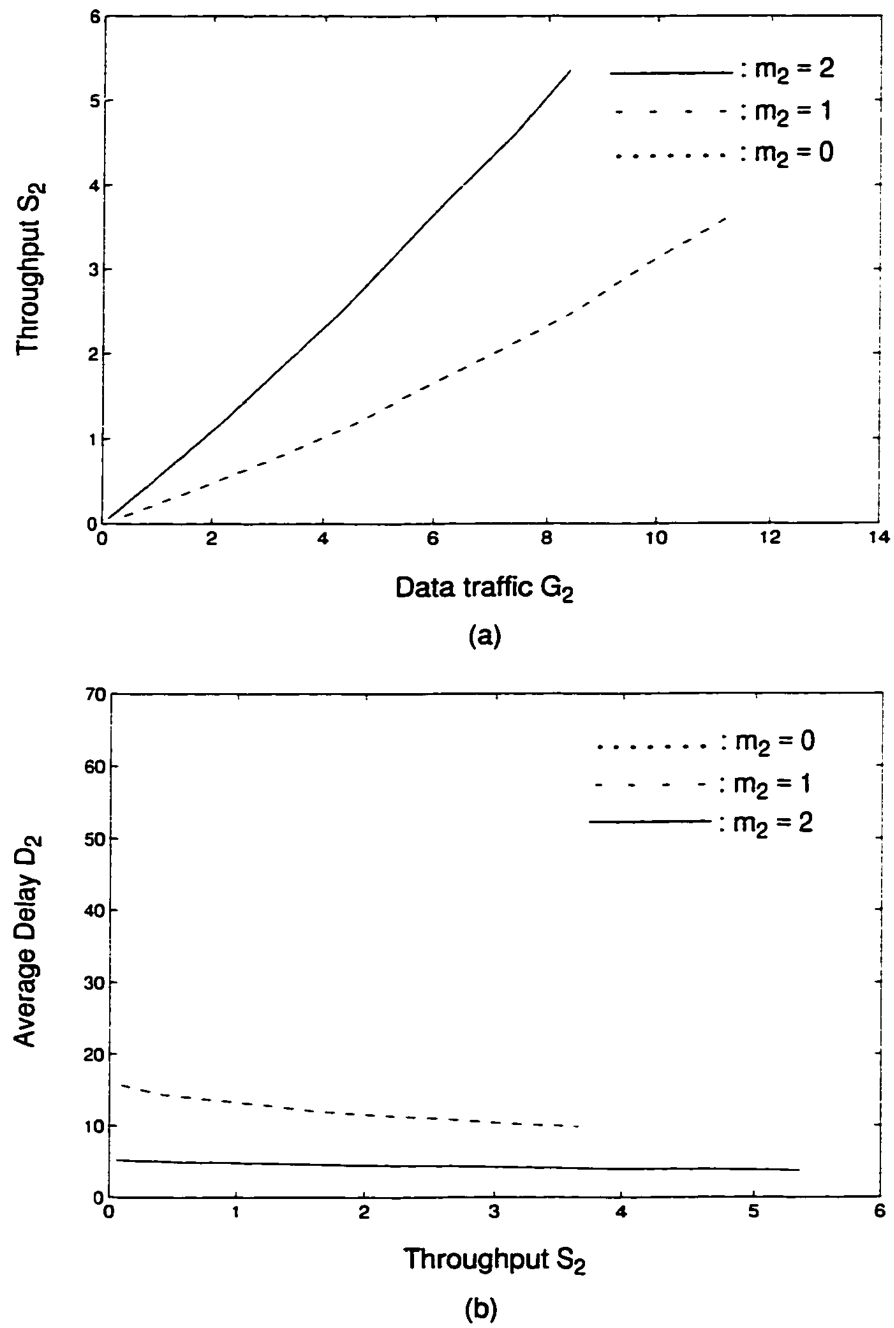


Fig. 3. 3 (a) Influence of data power level on throughput and offered data traffic. $m_1 = 0$ dB. $\sigma_1 = 1$ dB, $\sigma_2 = 1$ dB. (b) Influence of data power level on the relation between throughput and delay. $m_1 = 0$ dB. $E(\tau) = 40$ ms. $\sigma_1 = 1$ dB, $\sigma_2 = 1$ dB.

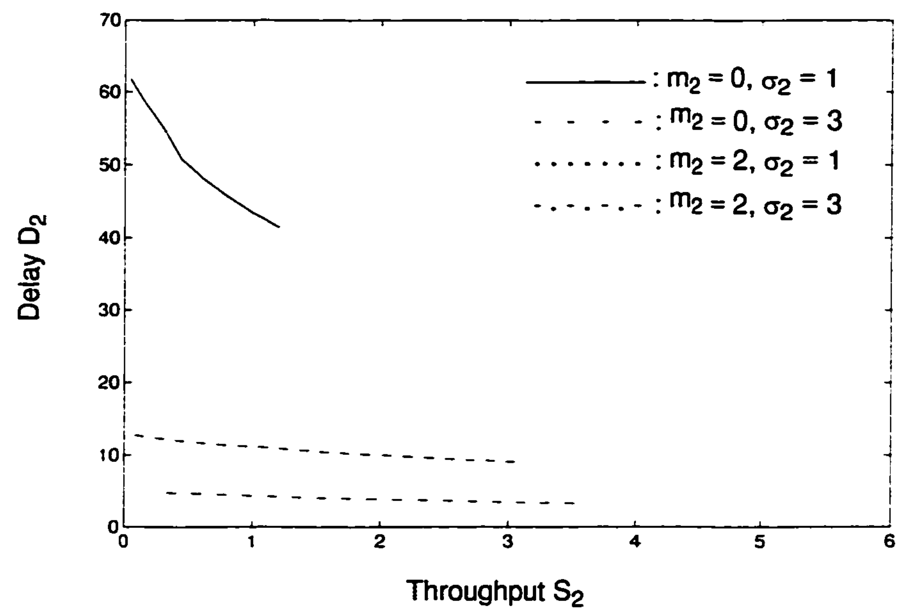


Fig. 3. 4 Delay as a function of throughput under good power control of voice service. $m_1 = 0$ dB, $\sigma_1 = 1$ dB. $E(\tau) = 40$ ms

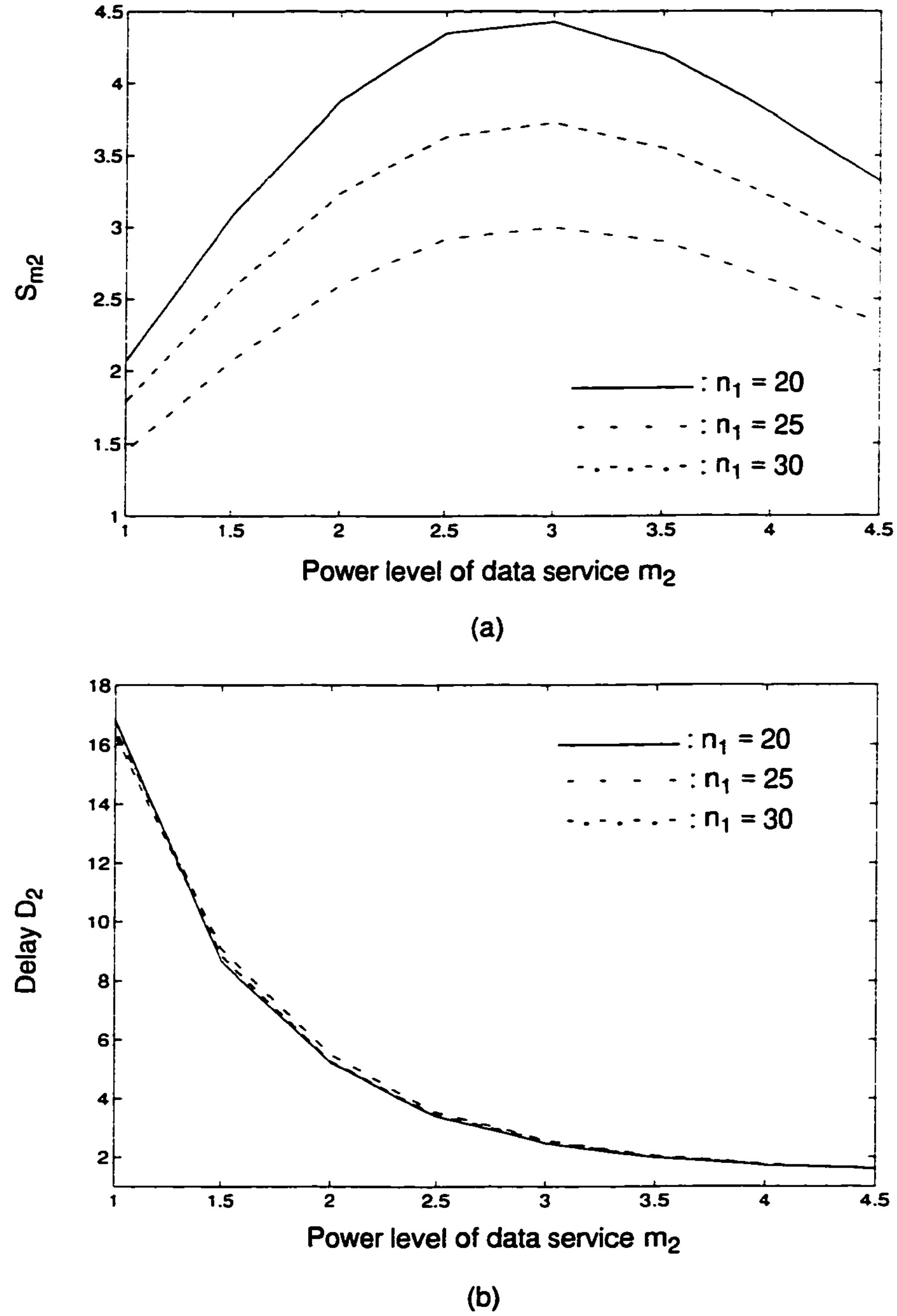


Fig. 3. 5 (a) Influence of m_2 on S_{m2}. $n_2 = 25$, $m_1 = 0$ dB, $\sigma_1 = 1$ dB, $\sigma_2 = 1$ dB. $PG_1 = 256$, $PG_2 = 256$. (b) Relation between data service power level and its delay. $\sigma_1 = 1$ dB, $\sigma_2 = 1$ dB, $m_1 = 0$ dB. $E(\tau) = 40$ ms

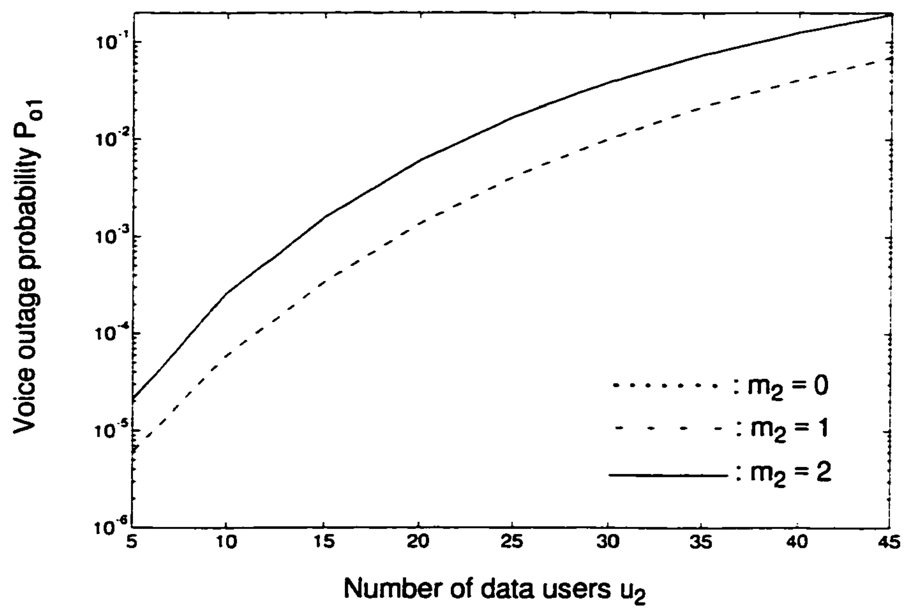


Fig. 3. 6 Voice outage probability. $u_1 = 20$, $m_1 = 0$ dB. $\sigma_1 = \sigma_2 = 1$ dB

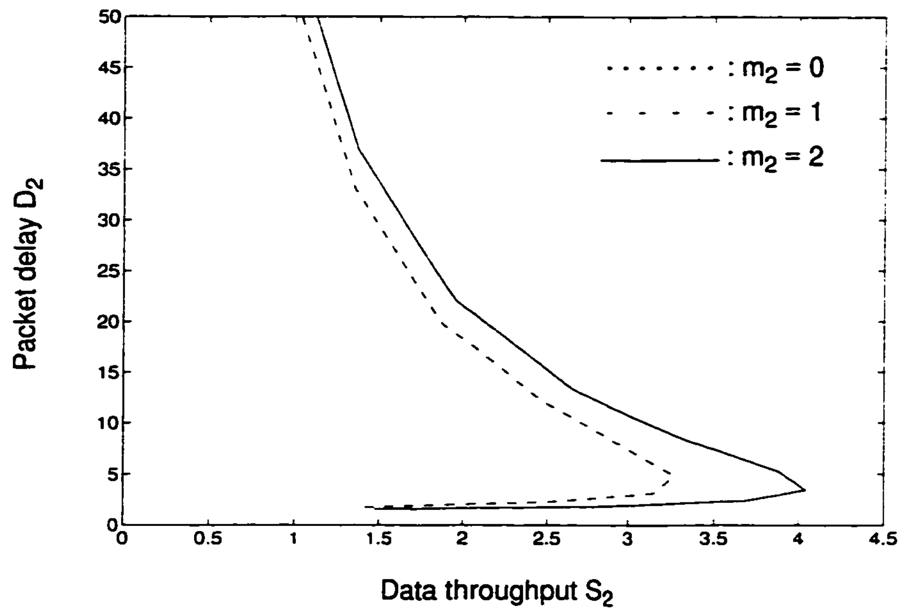
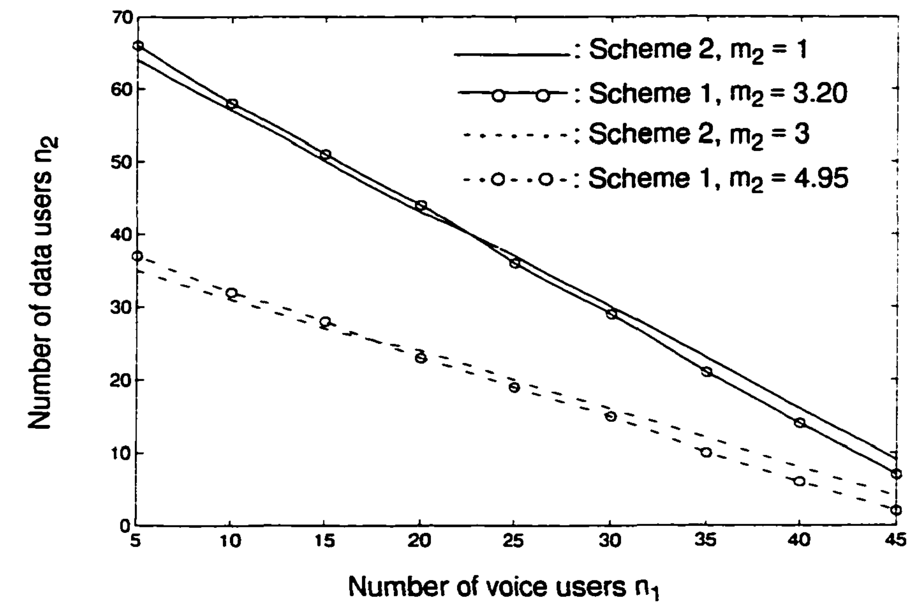
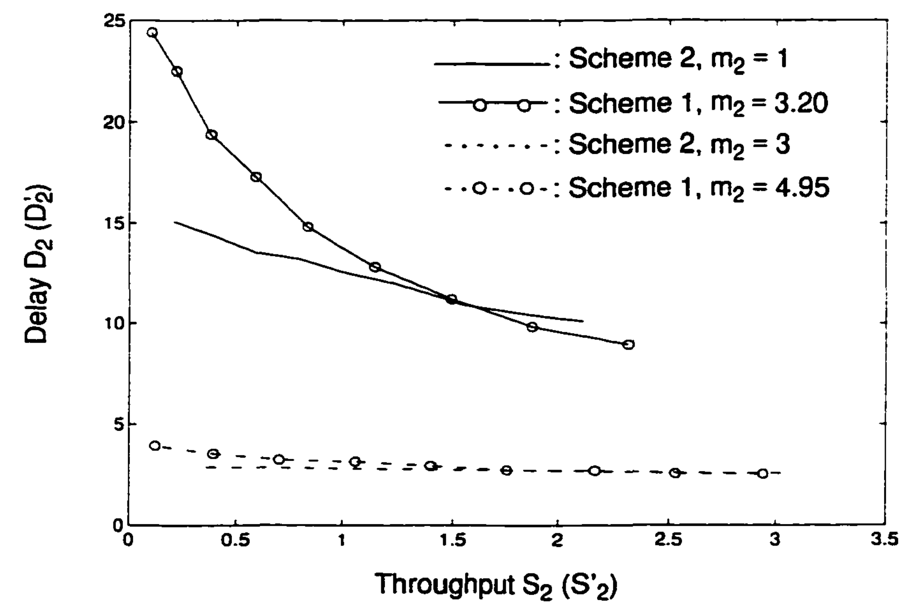


Fig. 3. 7 Packet delay of data users. $u_l = 20$, $m_l = 0$ dB, $m_2 = 2$ dB.



(a)



(b)

Fig. 3. 8 (a) Capacity of mixed traffic of two schemes. $m_1 = 0$ dB, $\sigma_1 = 1$ dB, $\sigma_2 = 1$ dB. In scheme 1, $PG_2 = 128$; in scheme 2, $PG_2 = 256$. (b) Comparison of throughput and delay between two schemes. $\sigma_1 = 1$ dB, $\sigma_2 = 1$ dB, $m_1 = 0$ dB. $E(\tau) = 40$ ms.

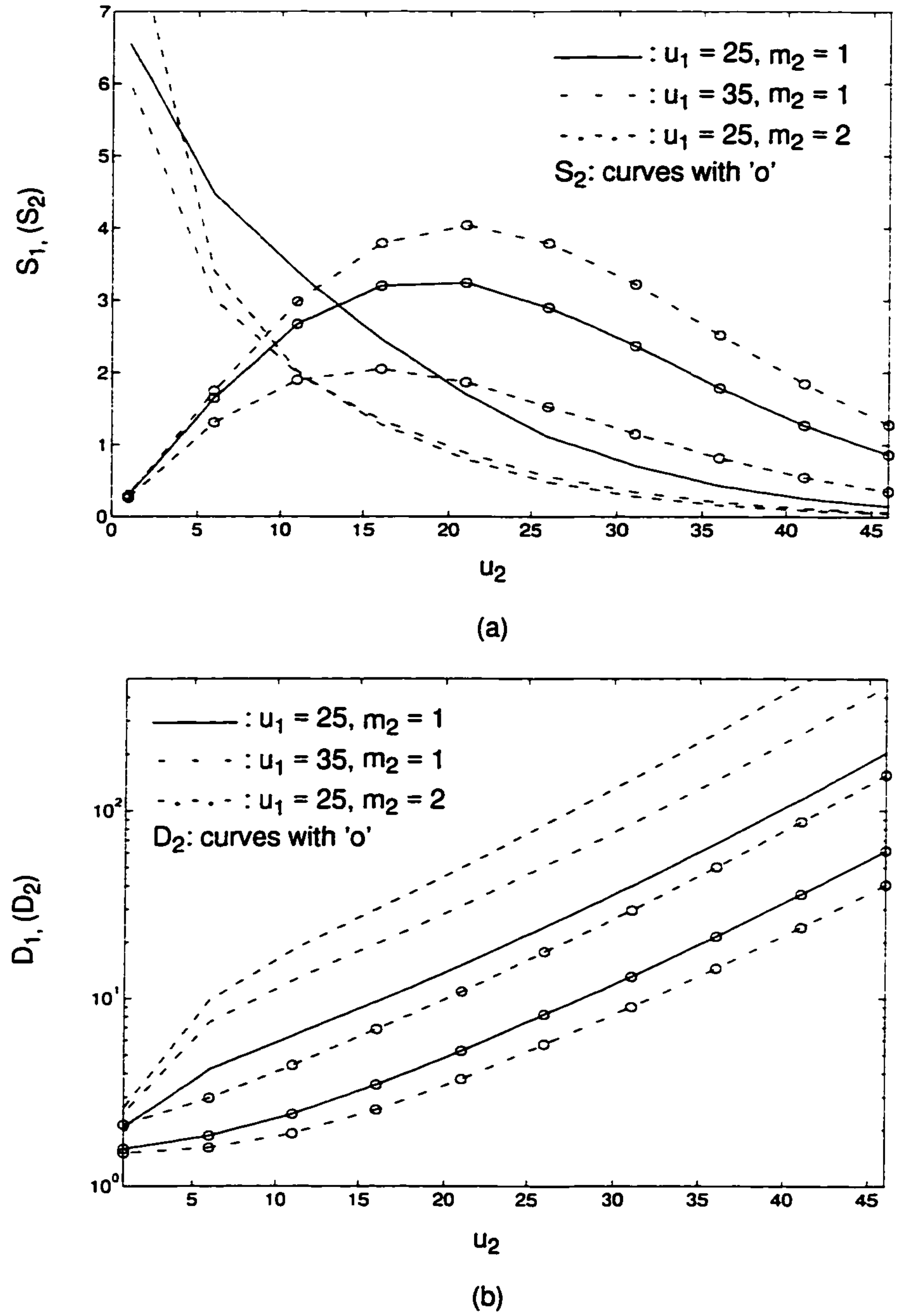


Fig. 3. 9 (a) Influence of u_2 on throughput. $u_2 = 1 \ldots 50$, $\alpha_1 = \alpha_2 = 0.3$, $m_1 = 0$ dB. (b) Influence of u_2 on packet delay. $u_2 = 1 \ldots 50$, $\alpha_1 = \alpha_2 = 0.3$, $m_1 = 0$ dB.

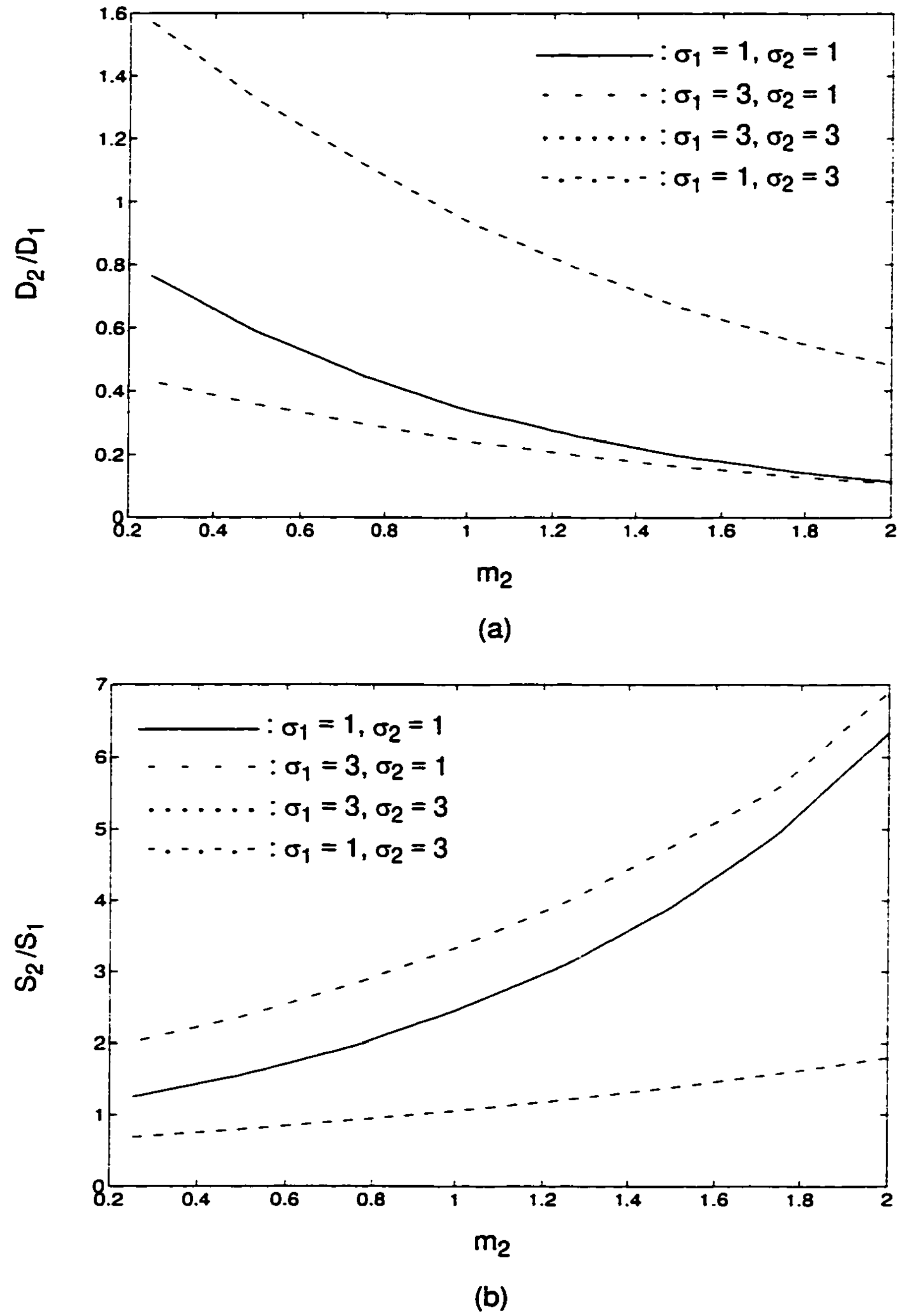


Fig. 3. 10 (a) Influence of m_2 on D_2/D_1. $u_1 = u_2 = 25$, $\alpha_1 = \alpha_2 = 0.3$, $m_1 =$ 0 dB. (b) Influence of m_2 on S_2/S_1. $u_1 = u_2 = 25$, $\alpha_1 = \alpha_2 = 0.3$, $m_1 = 0$ dB.

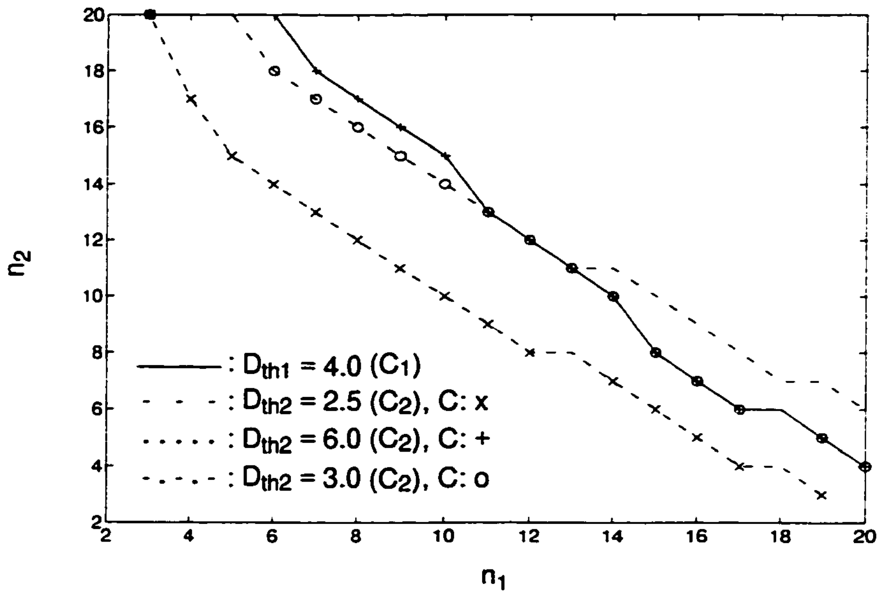


Fig. 3. 11 Fixed D_{th1} and variable D_{th2}, $m_1 = 0$ dB., m_2=1 dB, $\sigma_1 = \sigma_2$ =3 dB, C_1, C_2 and $C = \min(C_1, C_2)$. $\alpha_1 = \alpha_2 = 0.3$.

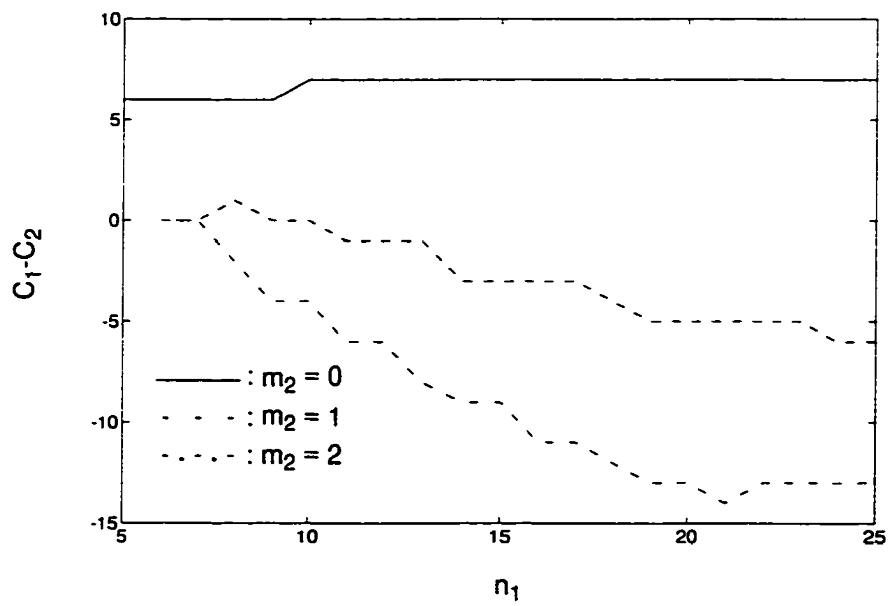


Fig. 3. 12 Influence of m_2 on C_1-C_2. $\alpha_1 = \alpha_2 = 0.3$, $m_1 = 0$ dB. $D_{th1} = 4$, $D_{th2} = 2.5$.

Chapter 4

Admission Control Policies in Multimedia Cellular Systems

4. 1 Channel Allocation

In multi-service cellular systems, there are Ω types of services. An MS can use any particular service. Each service has two types of traffic, new call traffic and handoff call traffic. Thus 2Ω different types of traffic should be considered when we design admission control policy (ACP). Due to the limited spectrum allocated to a wireless system and therefore to a cell, the number of (basic) channels, say N, in a cell is also limited. A basic channel is defined to have the minimum transmission capability, which can support a call of at least one service. More than one basic channel can be assigned to an MS at a time if a service requiring a higher transmission capability is in use. Assuming there is a set S_c of $N = \|S_c\|$ basic channels in a cell, we may allocate these basic channels to each and every service/call type, i. e. to determine a mapping between each type and the channel set accessible by that type. A user's service/call type determines the channel subset open to him. ACP runs according to channel allocation profile as well as other particular rules. Thus we include channel allocation as an integral part of ACP. Denote that for the type-i service calls, its accessible channel sets are S_{ni} and S_{hi} for its new calls and handoff calls respectively, then:

$$S_{ni} \subseteq S_c \text{ and } S_{hi} \subseteq S_c \text{, for } i = 1, 2, \ldots, \Omega. \tag{4.1}$$

We also denote that $S_i = S_{ni} \cup S_{hi}$ as the set of all basic channels allocated to the entire type-i service. The selection of S_i depends on the characteristics of the type-i service, such as its degree of importance, blocking probability requirement, and traffic load. Different channel sets accessible by different service types may overlap. So, the portion of the channel set space that overlaps is shared (or competed) by more than a single service type. The non-overlap part is allocated exclusively to a particular service. Since all channels have to be allocated, we always have:

$$S_c = S_{n1} \cup S_{h1} \cup S_{n2} \cup S_{h2} \cup \ldots S_{n\Omega} \cup S_{h\Omega}. \quad (4.2)$$

There are two extreme cases of channel allocation scheme. The first one is:

$$S_c = S_{n1} + S_{h1} + \ldots S_{n\Omega} + S_{h\Omega}. \quad (4.3)$$

In the above case, no channel in a cell is shared by any two types. The other case is:

$$S_c = S_{n1} = S_{h1} = \ldots S_{n\Omega} = S_{h\Omega}, \quad (4.4)$$

i. e. all channels in a cell are shared by all types as in the case of nonprioritized ACP described in Section 4. 3. The capacity of a cell is reached upon no free channel available:

$$C = C|_{\overline{S_c} = \theta}. \quad (4.5)$$

where θ denotes an empty set and $\overline{S_c}$ is the set of available free channels in a cell. When $\overline{S_c} = \theta$, access requests from all types will be blocked by BS. The cell capacity is represented by a set of vectors C with vectors in the form of:

$$\phi = (n_1, n_2, \ldots, n_\Omega), \text{ for } \phi \in C, \quad (4.6)$$

where n_i is the total number of the type-i service ongoing calls in the cell. All ϕ must satisfy Eq. (4.5).

Calls of different services, if admitted, may take different amount of system resources according to their QoS requirement and/or data rate [30]. An example is shown in Chapter 3. The capacity of multi-service (voice and data for example) system in Fig. 3. 2 (a) clearly indicates that different services require different amount of system resources (unless the capacity curve is in the from of $n_1 + n_2 = \text{Constant}$). If a type-$i$ call requires k_i basic channels to carry, the maximum number of type-i ongoing new calls and the maximum number of type-i ongoing handoff calls are given by:

$$N_{ni} = \lfloor \|S_{ni}\| / k_i \rfloor \text{ and } N_{hi} = \lfloor \|S_{hi}\| / k_i \rfloor, \text{ for } i = 1, 2, \ldots, \Omega, \tag{4.7}$$

respectively, where $\|S_{ni}\|$ ($\|S_{hi}\|$) is the total number of basic channels allocated to type-i new (handoff) calls. It is assumed that $k_1 = 1$ and $k_i \geq 1$ for $i > 1$. In this chapter, we assume that k_i is an integer. The assumption simplifies the derivation of equations. We further denote:

$$N_i = \lfloor \|S_{ni} \cup S_{hi}\| / k_i \rfloor, \tag{4.8}$$

as the maximum number of type-i ongoing calls. Therefore, $0 \leq u_i \leq N_i$ is the number of ongoing calls of type-i service.

There are two contradicting objectives to be considered in admission control design. One is to ensure fair access among services, which does not exist in a single service system. Here we assume that maximum fairness means minimum difference of blocking probabilities among all types. However, due to the nature of services/calls, we sometimes have also to protect certain type(s) of calls, such as handoffs. Therefore, call protection in a multi-service scenario becomes the second objective. Since protecting some types will affect the rest, it is necessary to analyze the performance of ACPs achieving those objectives. Generally, ACPs should treat different services differently based on certain criteria.

To simplify the analysis in this chapter, we let $\Omega = 2$ and assume that Service 1 and Service 2, share those basic channels according to certain mapping profile. Therefore, four types of call are considered: new calls of Service 1, handoff calls of Service 1, new calls of Service 2, and handoff calls of Service 2.

One of the simplest ACPs is to allocate certain number of channels to each service without any channel sharing between them. New calls and handoff calls of a service share all allocated channels of that service, i. e. $S_{n1} = S_{h1}$, $S_{n2} = S_{h2}$, and $S_c = S_{n1} + S_{n2}$. A coming call of a service is assigned channel(s) until all allocated channels of that service are occupied. This no-sharing allocation results in two independent subsystems. However, if the subsystem of one service is full while the other still has free channel(s), the channel(s) can not be used by calls of the former service. The efficiency of this ACP is not high [40]. Therefore, ACPs allowing the channel sharing are of our interest.

To address the cell behavior, the state of a cell is defined. When there are u_1 ongoing calls of Service 1 and u_2 ongoing calls of Service 2 in a cell, we say that the cell is in the state of (u_1, u_2). The capacity of the cell becomes *a set of integer pairs* $\{(n_1, n_2)\}$ such that when system is in the state of (n_1, n_2), no further accessing request in the cell will be accepted because the cell is full. In a DS/CDMA system, this is to keep the mutual interference below a certain threshold. In a TDMA system, it means that all time slots are occupied. Of course, C is a subset of $\{(u_1, u_2)\}$. We further denote $\{(n_1 - m_1, n_2 - m_2)\}$ as the set of all the states such that, if a cell's state is in this set, the cell will become full after its BS exactly accepting m_1 more Service 1 calls and m_2 more Service 2 calls (suppose no ongoing call ends during the accepting period).

We will investigate how different ACPs affect the system performance in terms of blocking probability, for each type of call considered, under a given amount of mixed traffic. The fair access issue will also be investigated. Furthermore, the impact of the distribution of traffic on the ACP performance will be given which is an issue not existing in single traffic systems. However, we do not discuss how to make the best trade-off among different ACPs, since it depends on particular system design requirements. A framework of analyzing admission control is built in this chapter.

4. 2 Mixed Traffic Model

We assume that the two call streams of the two services are two independent Poisson processes. For each service traffic we adopt the model of [17]. It is assumed that no two or more handoff/new calls arrive at the same time. The call arriving rates of Service 1 and 2 are (new call rate plus handoff call rate):

$$\begin{cases} \lambda_1 = \lambda_{n1} + \lambda_{h1} \\ \lambda_2 = \lambda_{n2} + \lambda_{h2} \end{cases} . \tag{4.9}$$

The number of MSs in a cell is much larger than N_1. The call arriving rate is not affected by the number of ongoing calls in the cell. The average service time of Service 1 and 2 are $1/\mu_{s1}$ and $1/\mu_{s2}$, respectively. The average residence time of MSs in a cell is $1/\mu_{r1} = 1/\mu_{r2}$, assuming that the mobility of both service users is the same and all cells are of the same size. Both service time and cell residence time are exponentially distributed. The average channel holding time of calls of the two services are exponentially distributed with mean $1/\mu_1 = 1/(\mu_{s1} + \mu_{r1})$ and $1/\mu_2 = 1/(\mu_{s2} + \mu_{r2})$, respectively. The traffic intensities are:

$$\begin{cases} \rho_1 = \lambda_1/\mu_1 \\ \rho_2 = \lambda_2/\mu_2 \end{cases} , \tag{4.10}$$

for Service 1 and 2, respectively. In addition, we also denote ρ_2/ρ_1 as a measure of traffic distribution between the two services, which is simply λ_2/λ_1 on condition that $\mu_1 = \mu_2$. When the blocking probability is far less than 1, we have $\lambda_{h1} \approx \mu_{r1}\lambda_{n1}/\mu_{s1}$ and $\lambda_{h2} \approx \mu_{r2}\lambda_{n2}/\mu_{s2}$ [17]. Then we can determine the call blocking probability of each type of calls under given N_1 as well as certain admission control rules. It is interesting to see how the blocking performance changes as the traffic or its distribution changes. Furthermore, the dissimilarity indicated by k also affects the blocking and fairness performance of ACPs as shown in the following sections.

4. 3 Nonprioritized Admission Control Policy

Nonprioritized means that access requests from all types of calls are treated equally by ACP when a BS makes channel assignment decision.

4. 3. 1 First-Come-First-Served policy

In this section, we always assume that one call of Service 2 is equivalent to $k_2 = k$ calls of Service 1, where k is an integer greater than 1. In other words, a call of Service 1 takes one basic channel, while a call of Service 2 takes k basic channels if accepted by a BS. Since the two services share all $N_1 = N$ basic channels, it is obvious that the two elements in the capacity data pair have the following relationship:

$$n_1 + kn_2 = N_1 . \tag{4.11}$$

It is also assumed that $N_2 = N_1/k$ is an integer. Either one or k basic channel(s) are assigned each time. When there are i ongoing Service 1 calls and j ongoing Service 2 calls, the number of available free basic channels is:

$$l_f = N_1 - i - kj . \tag{4.12}$$

This ACP's channel assignment follows a first-come-first-served (FCFS) rule [32]. The channel allocation is described as $S_c = S_{n1} = S_{h1} = S_{n2} = S_{h2}$. No matter what type of call, until the capacity limit is reached, i. e. *nonprioritized.* If the number of the free basic channels is less than k, any Service 2 access request will be rejected. However, a Service 1 access request is rejected only when there is no free basic channel in the cell. All blocked calls are cleared. The behavior of the cell can be described by a two dimensional Markov chain. Its state transition diagram is shown in Fig. 4. 1 with each circle being a state. (n_1, n_2) pairs are emphasized by thicker circles. There are $N_2 + 1$ such states that a cell is full and BS rejects all access requests from both services.

This is the basic ACP of sharing the entire N_1 basic channels by two services. Prioritized ACPs are based on this one.

We develop a two-dimensional Markov chain to model the cell behavior and use it to examine the performance of different ACPs. For the case of $k = 1$, a one-dimensional Markov chain is able to model the cell [41]. To get the blocking probabilities of both services, the steady state probability that a cell is in state (i,j), $q_{i,j}$ (where $0 \le i, 0 \le j, 0 \le i + kj \le N_1$), are needed. When a cell is at stable state, the following linear equations hold:

$$\begin{cases} (i\mu_1 + j\mu_2) q_{i,j} = \lambda_1 q_{i-1,j} + \lambda_2 q_{i,j-1} & (i,j) \in \{ (n_1, n_2) \} \\ (j\mu_2 + \lambda_1 + i\mu_1) q_{i,j} = (i+1)\mu_1 q_{i+1,j} + \lambda_1 q_{i-1,j} + \lambda_2 q_{i,j-1} & (i,j) \in \bigcup_{l=1}^{k-1} \{ (n_1 - l, n_2) \} \\ (\lambda_1 + \lambda_2 + i\mu_1 + j\mu_2) q_{i,j} = (j+1)\mu_2 q_{i,j+1} + \lambda_2 q_{i,j-1} & \\ \qquad + (i+1)\mu_1 q_{i+1,j} + \lambda_1 q_{i-1,j} & \text{else} \end{cases} \quad (4.13)$$

Special cases, such as $q_{0,0}$, are included by setting $q_{0,-1}$ and $q_{-1,0}$ to 0 in the third line of Eq. (4.13). In other words, $q_{i,j}$s with invalid subscript i, j in equations are treated as 0 throughout this chapter. By this way, we reduce the number of lines in equations. It may be verified by direct substitution that $q_{i,j} = q_{0,0}(\lambda_1/\mu_1)^i(\lambda_2/\mu_2)^j/(i!j!)$. Because $\sum q_{i,j} = 1$, we can obtain $q_{0,0}$ as:

$$q_{0,0} = \left[\sum_{u_2=0}^{N_2} \sum_{u_1=0}^{k(N_2-u_2)} \left(\frac{\lambda_1}{\mu_1}\right)^{u_1} \left(\frac{\lambda_2}{\mu_2}\right)^{u_2} \left(\frac{1}{u_1! u_2!}\right) \right]^{-1}. \quad (4.14)$$

Therefore,

$$q_{i,j} = \left(\frac{\lambda_1}{\mu_1}\right)^i \left(\frac{\lambda_2}{\mu_2}\right)^j \left(\frac{1}{i!j!}\right) q_{0,0}. \quad (4.15)$$

Since no more calls of Service 1 can be admitted, when cell is in any states of $\{ (n_1, n_2) \}$, the blocking probability of Service 1, for both new calls and handoff calls, is:

$$P_1^{(n)} = \sum_{j=0}^{N_2} q_{i,j} = \sum_{j=0}^{N_2} \left(\frac{\lambda_1}{\mu_1}\right)^i \left(\frac{\lambda_2}{\mu_2}\right)^j \left(\frac{1}{i!j!}\right) q_{0,0}, \quad (4.16)$$

where $i = N_1 - kj$. However, no more calls of Service 2 can be admitted when system is in states of (n_1, n_2) or $(n_1 - l, n_2)$ ($1 \le l \le k-1$). The blocking probability of all Service 2 calls is:

$$P_2^{(n)} = \sum_{j=0}^{N_2} \sum_{i=i_o}^{N_1-kj} \left(\frac{\lambda_1}{\mu_1}\right)^i \left(\frac{\lambda_2}{\mu_2}\right)^j \left(\frac{1}{i!j!}\right) q_{0,0}, \quad (4.17)$$

where $i_o = max(0, N_1 - kj - k + 1)$.

Therefore, it results in $P_1^{(n)} < P_2^{(n)}$. In general, calls requiring more system resource (larger channels) suffer higher blocking probability if no priority is given to them. The difference between $P_2^{(n)}$ and $P_1^{(n)}$ increases as the k increases.

As an indicator of fairness in multi-service systems, blocking probabilities of different type of calls can be normalized to the highest blocking rate. In this case, the fairness factor is defined as:

$$r_f^{(n)} = \frac{P_1^{(n)}}{P_2^{(n)}}. \tag{4.18}$$

Therefore, $0 < r_f^{(n)} \le 1$. $r_f^{(n)} = 1$ corresponds to the perfect fairness. The fairness loses as $r_f^{(n)}$ decreases. $r_f^{(n)}$ is affected by the ACP, traffic density and traffic distribution as can be seen later.

4. 4 Prioritized Admission Control Policies

In a multi-service network, different types of call may be considered to have different degree of importance. The blocking of more important calls is less desirable and more costly than the blocking of less important ones. The degree of importance can be set based on certain criteria. For example, within a service, handoff calls are more important than new calls since forced terminations is more uncomfortable than blocking new calls. Between services, high rate service are important than low rate service, etc. Before designing and examining an ACP, we should know the importance of each type. In this chapter, we assume the following sequence: Service 2 handoff, Service 2 new call, Service 1 handoff and Service 1 new call, in the order of decreasing importance.

Since Service 2 handoff calls are assumed to have the highest degree of importance, special protection should be given to them by using *prioritized* ACPs. Prioritization can be in different ways. One is to put more access limitations on nonprioritized types by channel allocation. The other is to reducing access limitations on prioritized type, such as the proposed call dropping ACP. Prioritization scarifies fairness in order to reduce the blocking rate of certain type(s) of call. The losing of fairness can be controlled by proper trade-offs. In multi-service scenario, prioritized ACPs are usually necessary.

In prioritized ACPs, all types of call can access channels by an FCFS rule until certain conditions are met. Then, prioritization arrangement begins to work.

4. 4. 1 Prioritized ACPs for service quality (I)

Access limitation on types with lower degree of importance

This is the case of sacrificing fairness for better service quality. To protect handoffs, the fairness becomes less important. Because Service 1 is already less blocked and Service 2 is assumed to be more important, we consider giving Service 2 handoffs priorities. This can be seen as an effort to protect a call already taking k basic channels (and its termination is considered to be more costly than the termination of a call taking only one basic channel). With four types of call considered, giving priority to a type means that more channels are allocated to that type. Access requests other than that type are rejected under certain conditions even if there is still room for them. Three cases with $k = 2$ are examined here.

- **Case 1**: only new calls of Service 2 will be blocked when cell is in any of the states of $(n_1, n_2 - 1)$ and $(n_1 - 1, n_2 - 1)$. The rest part of the Markov chain of this case is the same as the FCFS ACP, i.e. following an FCFS rule.
- **Case 2**: all Service 1 inquiries are rejected when system is in any of the states of $(n_1, n_2 - 1)$. The rest part of the ACP is the same as that of Case 1.
- **Case 3**: no Service 1 call will be admitted if system at the states of $(n_1, n_2 - 1)$ or $(n_1, n_2 - 2)$. The rest part of the ACP is the same as that of Case 2.

To limit Service 2 new call access is to prevent further congestion under heavy traffic in these cases. In summary, $\|S_{h2}\| = N_1$ and $\|S_{n2}\| = N_1 - k$ in all cases; $\|S_{n1}\| = \|S_{h1}\| = N_1$, $N_1 - 1$, and $N_1 - 2$ in Case 1, 2, and 3, respectively. The Markov chains of these three ACPs are shown in Fig. 4. 8 (a), (b), and (c) respectively. The transition probabilities among some states are Case dependent as shown. In Case 1, the following equations hold:

$$\begin{cases} (n_1\mu_1 + n_2\mu_2)\, q_{n_1, n_2} = \lambda_{h2} q_{n_1, n_2 - 1} + \lambda_1 q_{n_1 - 1, n_2} \\ [n_2\mu_2 + (n_1 - 1)\,\mu_1 + \lambda_1]\, q_{n_1 - 1, n_2} = \lambda_{h2} q_{n_1 - 1, n_2 - 1} + n_1\mu_1 q_{n_1, n_2} + \lambda_1 q_{n_1 - 2, n_2} \\ (\lambda_{h2} + \lambda_1 + i\mu_1 + j\mu_2)\, q_{i,j} = (j+1)\,\mu_2 q_{i,j+1} + \lambda_2 q_{i,j-1} + (i+1)\,\mu_1 q_{i+1,j} + \lambda_1 q_{i-1,j} \end{cases} . \quad (4.19)$$

The subscripts in the third line of Eq. (4.19) must be $(i,j) \in \{(n_1, n_2 - 1) \cup (n_1 - 1, n_2 - 1)\}$. For other $q_{i,j}$, Eq. (4.13) still holds. For Service 1, its blocking probability $P_1^{(p1)}$, no matter new calls or handoffs, is given by Eq. (4.16). For Service 2, its blocking probability of handoff calls is still given by Eq. (4.17) as $P_{h2}^{(p1)} = P_2^{(n)}$, i. e. by the same right-hand expression. The blocking probability of new calls becomes:

$$P_{n2}^{(p1)} = \sum_{all\,(i,j) \in C_{(1)}} q_{i,j}, \quad (4.20)$$

where $C_{(1)} = \{(n_1 - 1, n_2 - 1) \cup (n_1 - 1, n_2) \cup (n_1, n_2 - 1) \cup (n_1, n_2)\}$.

For Case 2, the equations in Eq. (4.21) can be derived according to Fig. 4. 8 (b) as:

$$\begin{cases} (n_1\mu_1 + n_2\mu_2)\, q_{n_1, n_2} = \lambda_{h2} q_{n_1, n_2 - 1} \\ [n_2\mu_2 + (n_1 - 1)\,\mu_1]\, q_{n_1 - 1, n_2} = \lambda_{h2} q_{n_1 - 1, n_2 - 1} + n_1\mu_1 q_{n_1, n_2} + \lambda_1 q_{n_1 - 2, n_2} \\ (\lambda_{h2} + \lambda_1 + i\mu_1 + j\mu_2)\, q_{i,j} = (j+1)\,\mu_2 q_{i,j+1} + \lambda_2 q_{i,j-1} + (i+1)\,\mu_1 q_{i+1,j} + \lambda_1 q_{i-1,j} \end{cases} . \quad (4.21)$$

The (i,j) in the third line of Eq. (4.21) is $(i,j) \in \{(n_1, n_2 - 1) \cup (n_1 - 1, n_2 - 1)\}$. Eq. (4.20) is still used to get Service 2 new call blocking probability in Case 2. Eq. (4.17) gives the Service 2 handoff call blocking probability. However, for Service 1 calls, their blocking probability is the summation of $q_{i,j}$ over $C_{(2)} = \{(n_1 - 1, n_2) \cup (n_1, n_2)\}$.

For Case 3, we have:

$$\begin{cases} (n_1\mu_1 + n_2\mu_2)\, q_{n_1, n_2} = \lambda_{h2} q_{n_1, n_2 - 1} \\ [n_2\mu_2 + (n_1 - 1)\,\mu_1]\, q_{n_1 - 1, n_2} = \lambda_{h2} q_{n_1 - 1, n_2 - 1} + n_1\mu_1 q_{n_1, n_2} \\ [\lambda_{h2} + (j-1)\,\mu_2 + i\mu_1]\, q_{i,j-1} = \lambda_2 q_{i,j-2} + (i+1)\,\mu_1 q_{i+1,j-1} + \lambda_1 q_{i-1,j-1} + j\mu_2 q_{i,j} \\ (\lambda_{h2} + \lambda_1 + i\mu_1 + j\mu_2)\, q_{i,j} = (j+1)\,\mu_2 q_{i,j+1} + \lambda_2 q_{i,j-1} + (i+1)\,\mu_1 q_{i+1,j} + \lambda_1 q_{i-1,j} \end{cases} . \quad (4.22)$$

In the third line of Eq. (4.22), $i = n_1$ and $j = n_2$. In the forth line of Eq. (4.22), $i = n_1 - 1$ and $j = n_2 - 1$. Eq. (4.20) and (4.17) are still used to get Service 2 new call and handoff blocking probabilities in Case 3, respectively. However, the blocking rate of Service 1 is the summation of $q_{i,j}$, over $C_{(3)} = \{ (n_1-2, n_2) \cup (n_1-1, n_2) \cup (n_1, n_2) \}$.

The fairness factors for the three cases are defined as:

$$r_{f1}^{(p1)} = \frac{P_1^{(p1)}}{P_{n2}^{(p1)}}, \text{ and } r_{f2}^{(p1)} = \frac{P_{h2}^{(p1)}}{P_{n2}^{(p1)}}. \tag{4.23}$$

Channel reservation scheme

To reserve channels in a two traffic scenario is different from that in a homogenous traffic scenario. The number of reserved channels is a multiple of k when reservation is for Service 2. Basically, in both scenario, a BS begins to reject access requirements from certain types of call when the number of remaining free basic channels in a cell is less than a preset value while the prioritized type can still be assigned channels. However, for a two service network, rejections at BSs begin to happen in a set of states instead of only one state. In addition, reservation affects the blocking probability of all types of call and may reduce the traffic load carried under given blocking requirement. Therefore, the cost on nonprioritized service has to be investigated.

To reserve channels for Service 2 handoffs, the following admission rules are assumed:

- the number of basic channels reserved *only* for Service 2 handoff calls is km,
- when the number of free basic channels is no greater than km, all access requests other than Service 2 handoffs are rejected,
- when the number of free basic channels is no less than $(k+1)m$, the FCFS rule is always followed by the ACP,
- when the number of free basic channels is less than $(k+1)m$ but greater than km, the FCFS rule is only followed by the ACP when accepting access requests other than Service 2 new calls; Service 2 new calls are rejected.

Thus there are km basic channels reserved for Service 2 handoff calls only as shown in

Fig. 4. 2. The channel allocation is: $\|S_{n1}\| = \|S_{h1}\| = N_1 - km$, $\|S_{h2}\| = N_1 - km$ and $\|S_{h2}\| = N_1$. In this case, the linear equations of those states affected by the reservation are:

$$\begin{cases} (i\mu_1 + j\mu_2) q_{i,j} = \lambda_{h2} q_{i,j-1} & (i,j) \in \{ (n_1, n_2) \} \\ (i\mu_1 + j\mu_2) q_{i,j} = \lambda_{h2} q_{i,j-1} + (i+1)\mu_1 q_{i+1,j} & (i,j) \in \bigcup_{l=1}^{k-1} \{ (n_1, n_2 - l) \} \\ (\lambda_{h2} + i\mu_1 + j\mu_2) q_{i,j} = \lambda_{h2} q_{i,j-1} + & (i,j) \in \bigcup \{ (l_1, l_2) \} \text{ where} \\ (j+1)\mu_2 q_{i,j+1} + (i+1)\mu_1 q_{i+1,j} & N_1 - km < l_1 + kl_2 < N_1 - k + 1 \\ (\lambda_{h2} + i\mu_1 + j\mu_2) q_{i,j} = \lambda_2 q_{i,j-1} + \lambda_1 q_{i-1,j} & (i,j) \in \bigcup \{ (l_1, l_2) \} \text{ where} \\ (j+1)\mu_2 q_{i,j+1} + (i+1)\mu_1 q_{i+1,j} & N_1 - km - k + 1 \le l_1 + kl_2 \le N_1 - km \\ (\lambda_1 + \lambda_{h2} + i\mu_1 + j\mu_2) q_{i,j} = \lambda_2 q_{i,j-1} + \lambda_1 q_{i-1,j} & (i,j) \in \bigcup \{ (l_1, l_2) \} \text{ where} \\ (j+1)\mu_2 q_{i,j+1} + (i+1)\mu_1 q_{i+1,j} & N_1 - km - k + 1 \le l_1 + kl_2 \le N_1 - km \end{cases} \quad (4.24)$$

The total number of states of a cell is given by: $n_s = (N_1/k + 1)^2 - m^2 - m$. The states affected are shown in Fig. 4. 2 (b). The blocking probabilities for Service 2 handoff calls, Service 2 new calls and all Service 1 calls are as follows:

$$P_{h2}^{(p2)} = \sum_{C_{(1)}} q_{i,j}, \text{ for } C_{(1)} = \bigcup \{ (l_1, l_2) \}, \quad (4.25)$$

where $N_1 - k + 1 \le l_1 + kl_2 \le N_1$.

$$P_{n2}^{(p2)} = \sum_{C_{(2)}} q_{i,j}, \text{ for } C_{(2)} = \bigcup \{ (l_1, l_2) \}, \quad (4.26)$$

where $N_1 - km - k + 1 \le l_1 + kl_2 \le N_1 - km$.

$$P_1^{(p2)} = \sum_{C_{(3)}} q_{i,j}, \text{ for } C_{(3)} = \bigcup \{ (l_1, l_2) \}, \quad (4.27)$$

where $N_1 - km \le l_1 + kl_2$.

The fairness factors are:

$$r_{f1}^{(p2)} = \frac{P_1^{(p2)}}{P_{n2}^{(p2)}}, \text{ and } r_{f2}^{(p2)} = \frac{P_{h2}^{(p2)}}{P_{n2}^{(p2)}}. \quad (4.28)$$

We expect that $r_{f2}^{(p2)} \to 0$ which reflects two facts: handoff calls are well protected and the fairness between Service 2 handoffs and the rest of calls is lost.

Similarly, if we change all λ_{h2} in this subsection by λ_2, this ACP gives priority to all Service 2 accesses. In this case, Eq. (4.25) is used to calculate the blocking probability of Service 2 $P_2^{(p2)} = P_{h2}^{(p2)} = P_{n2}^{(p2)}$. For all calls of Service 1, Eq. (4.27) is still valid.

4. 4. 2 Prioritized ACP for service quality (II)

Access enhancement scheme

When a handoff call of Service 2 requires basic channels, one of the scenarios is that the number of free basic channels in the target cell is less than k. In this case, we may either temporarily reduce the transmission rate of that handoff call or be ready to suffer quality degradation, just like using soft capacity in a CDMA system, to avoid the denial of admission. However, the feasibility of the above schemes depends on the nature of Service 2. For example, if the service is sensitive to delay and quality degradation, other scheme is necessary. In order to make room for the handoff, dropping of ongoing call(s) of Service 1 is an alternative scheme. In this case, the access capability of Service 2 handoff is further enhanced and the FCFS rule is not always followed. By doing so, bandwidth utilization and the average traffic carried in a cell are increased. This policy is more attractive when the difference between the two services, in terms of k and/or the degree of importance, is larger. This ACP removes some access limitations on Service 2 handoffs. However, the increase of Service 1 dropping is needed to be addressed to see the cost. We assume that:

- the number of ongoing Service 1 call(s) that can be dropped is limited to a threshold k',
- when a cell is full, all access requests including Service 2 handoffs are rejected,
- the FCFS rule is always followed unless a Service 2 handoff call arrives and the number of free basic channels is no less than $k-k'$. If this happens, ongoing call(s) is dropped to make *just enough room* for the coming handoff request.

The channel allocation of this ACP is the same as the FCFS ACP. The Markov model of this ACP, which is different, is shown in Fig. 4. 3. The part same as that of the FCFS

ACP in Fig. 4. 1 is not shown for simplicity. For these states affected by droppings we have:

$$\begin{cases} (i\mu_1 + j\mu_2)\, q_{i,j} = \lambda_2 q_{i,j-1} + \lambda_1 q_{i-1,j} + \sum_{l=1}^{k'} \lambda_{h2} q_{i+l,j-1} & (i,j) \in \{ (n_1, n_2) \} \\ (\lambda_1 + i\mu_1 + j\mu_2 + \lambda_{h2})\, q_{i,j} = \lambda_2 q_{i,j-1} + \lambda_1 q_{i-1,j} + (i+1)\mu_1 q_{i+1,j} & (i,j) \in \bigcup_{l=1}^{k'} \{ (n_1 - k + l, n_2) \} \end{cases} \qquad (4.29)$$

For the remaining states, Eq. (4.13) still holds. The blocking probability of Service 2 new calls and Service 2 handoff calls are given, respectively, by:

$$P_{n2}^{(p3)} = \sum_{\{(i,j)\}} q_{i,j}, \text{ for } (i,j) \in \bigcup_{u_1=0}^{k-1} \{ (n_1 - u_1, n_2) \} , \qquad (4.30)$$

$$P_{h2}^{(p3)} = \sum_{\{(i,j)\}} q_{i,j}\,. \text{ for } (i,j) \in \bigcup_{u_1=0}^{k-k'-1} \{ (n_1 - u_1, n_2) \} . \qquad (4.31)$$

To evaluate the Service 1 call dropping probability in the prioritized ACP, the state of (n_1, n_2) is divided into $k' + 1$ parts. The first part is due to accepting one call without dropping and the $l + 1$ -th part is due to accepting one Service 2 handoff call and dropping l ongoing Service 1 calls, as shown in Fig. 4. 3 (b). Denoting the probabilities of those parts as $q_{n_1,n_2}^{(0)}$ and $q_{n_1,n_2}^{(l)}$, we have:

$$(n_1\mu_1 + n_2\mu_2)\, q_{n_1,n_2}^{(0)} = \lambda_2 q_{n_1,n_2-1} + \lambda_1 q_{n_1-1,n_2} , \qquad (4.32)$$

$$(n_1\mu_1 + n_2\mu_2)\, q_{n_1,n_2}^{(l)} = \lambda_{h2} q_{n_1+l,n_2-1} \quad , \text{for } l = 1, 2, \ldots, k' . \qquad (4.33)$$

Since the probability of state (n_1, n_2) , q_{n_1,n_2}, is the sum of $q_{n_1,n_2}^{(0)}$ and all $q_{n_1,n_2}^{(l)}$, we have q_{n_1,n_2} as:

$$q_{n_1,n_2} = \sum_{l=0}^{k'} q_{n_1,n_2}^{(l)} = \frac{\lambda_2 q_{n_1,n_2-1} + \lambda_1 q_{n_1-1,n_2}}{n_1\mu_1 + n_2\mu_2} + \sum_{l=1}^{k'} \frac{\lambda_{h2} q_{n_1+l,n_2-1}}{n_1\mu_1 + n_2\mu_2} . \qquad (4.34)$$

Since the second term on the right-hand of Eq. (4.34) is due to droppings, it should be eliminated when calculating Service 1 new call blocking probability $P_{n1}^{(p3)}$. The dropping probability of Service 1 ongoing calls is:

$$P_{d1}^{(p3)} = \sum_{\{(n_1, n_2)\}} \sum_{l=1}^{k'} q_{i,j}^{(l)} \quad . \tag{4.35}$$

The average number of calls dropped simultaneously for a Service 2 handoff call is:

$$n_{d1}^{(p3)} = \sum_{\{(n_1, n_2)\}} \frac{1}{q_{i,j} - q_{i,j}^{(0)}} \sum_{l=1}^{k'} l q_{i,j}^{(l)} \, . \tag{4.36}$$

By denoting the probability of dropping l call(s) for a Service 2 handoff call as $P_{d,l}^{(p3)}$, we have:

$$P_{d,l}^{(p3)} = \sum_{\{(n_1, n_2)\}} q_{i,j}^{(l)} \text{, for } l = 1, 2, \ldots, k' . \tag{4.37}$$

Although the dropping of Service 1 calls is not due to handoff failures, it is the same as handoff failure in the sense that calls are terminated prematurely. We thus combine the handoff failures and the droppings and call it by handoff blocking probability in general. Remembering that dropping is part of it, we find this blocking probability becoming:

$$P_{h1}^{(p3)} = \sum_{\{(n_1, n_2)\}} q_{i,j} \, . \tag{4.38}$$

Then the Service 1 new call blocking probability is the summation of over $\{(n_1, n_2)\}$ but removing the dropping probability:

$$P_{n1}^{(p3)} = \sum_{\{(n_1, n_2)\}} q_{i,j} - P_{d1}^{(p3)} \quad . \tag{4.39}$$

To determine a suitable k', denoted as k_t, we introduce a cost factor $1 < \theta_c < k$. θ_c indicates that it is considered worth to drop, in average, θ_c ongoing Service 1 calls to save one Service 2 handoff call but not more than that. θ_c may not necessarily be an integer.

Therefore,

$$k_t = max(k'|_{n_{d1}^{(p3)} \le \theta_c}) , \tag{4.40}$$

under a given amount of traffic. Due to droppings, all blocking probabilities involved are changed compared with the FCFS case. They will be shown by numeral examples in the next section. Since the blocking probability of Service 2 new calls is the highest, the fairness factors are:

$$r_{f1}^{(p3)} = \frac{P_{n1}^{(p3)}}{P_{n2}^{(p3)}} , \quad r_{f2}^{(p3)} = \frac{P_{h1}^{(p3)}}{P_{n2}^{(p3)}} \text{ and } r_{f3}^{(p3)} = \frac{P_{h2}^{(p3)}}{P_{n2}^{(p3)}} . \tag{4.41}$$

With the increase of k', the handoff blocking of Service 2 decreases and the Service 1 blocking increases. Also notice that the nonprioritized ACP is a special case of this ACP with $k' = 0$. After dropping call(s), a cell becomes full. Further dropping only happens when there is no enough room for an arriving Service 2 handoff again and the number of call to be dropped for that handoff is no greater than k'.

If we extend this priority from Service 2 handoff calls to all Service 2 calls, the blocking performance of the ACP can be obtained by substituting all λ_{h2} by λ_2 in this subsection. The blocking probability of all Service 2 calls, denoted as $P_2^{(p3)}$, is obtained by Eq. (4.31) and there is no difference in the treatment of Service 2 new calls and handoff calls. Eq. (4.38) and (4.39) are still valid for Service 1.

How to select Service 1 calls that are going to be dropped is another interesting topic. Some possible schemes are:

- randomly select calls to drop,
- terminate calls coming from the MSs near cell boundary (a termination may also be followed by a handoff procedure to neighboring cell if possible, i. e. a forced handoff).

The first scheme is easy to implement. The second has the advantage of reducing call dropping rate, but needs certain overhead to make the selection. In present analysis, we do not consider forced handoffs and assume randomly dropped Service 1 calls being lost.

4. 4. 3 Prioritized ACP for fairness

Soft capacity is a feature of DS/CDMA systems. It is possible to break a capacity limitation to allow certain service calls being admitted as a way of giving priority. We consider the case of $k = 2$. To solve the problem of $P_1^{(n)} < P_2^{(n)}$, we may allow a Service 2 call to access the system even if the system is already in one of the states $(n_1 - 1, n_2)$ such that: $\|S_{n1}\| = \|S_{h1}\| = N_1$ and $\|S_{n2}\| = \|S_{h2}\| = N_1 + 1$. In other words, Service 2 has its priority. The cost is temporary quality degradation due to more frequent outage and/or longer delay at new states of $(n_1 - 1, n_2 + 1)$. Hence, the soft capacity feature of DS/CDMA network is used here. The Markov chain is changed from the one of the FCFS ACP by adding some new states $(n_1 - 1, n_2 + 1)$. The transition probabilities among the states $(n_1 - 1, n_2)$, $(n_1 - 1, n_2 + 1)$ and $(n_1 - 2, n_2 + 1)$ are shown in Fig. 4. 20. The remaining part of the chain is unchanged. Equations for the new states and their neighbors are:

$$\begin{cases} [(n_1 - 1)\mu_1 + (n_2 + 1)\mu_2]\, q_{n_1 - 1, n_2 + 1} = \lambda_2 q_{n_1 - 1, n_2} \\ [(n_1 - 2)\mu_1 + (n_2 + 1)\mu_2]\, q_{n_1 - 2, n_2 + 1} = \lambda_1 q_{n_1 - 3, n_2 + 1} + \lambda_2 q_{n_1 - 2, n_2} + (n_1 - 1)\mu_1 q_{n_1 - 1, n_2 + 1} \end{cases} \quad .(4.42)$$

For remaining $q_{i,j}$, Eq. (4.13) still holds. Then the probability that the system uses its soft capacity is:

$$P_{sc}^{(p4)} = \sum_{all(i+1, j-1) \in C} q_{i,j}. \quad (4.43)$$

The corresponding blocking probabilities for both services are equal and given by:

$$P_1^{(p4)} = P_2^{(p4)} = \sum_{all(i,j) \in C} q_{i,j} + P_{sc}^{(p4)}. \quad (4.44)$$

Finally, the fairness factor is:

$$r_f^{(p4)} = \frac{P_{n1}^{(p4)}}{P_{n2}^{(p4)}} \equiv 1. \quad (4.45)$$

Performance degradation can be calculated by Eq. (3.13) and (3.16) in the case that a slotted ALOHA access, as the one stated in Chapter 3, is used in the reverse link.

Because no closed form is found for the above prioritized ACPs' $q_{i,j}$, we order all cell states of a particular ACP lexicographically to build a set of linear equations in the form of $\boldsymbol{AQ} = \boldsymbol{B}$, where $\boldsymbol{A}$ is coefficient matrix, $\boldsymbol{Q}$ is vector of all states, and $\boldsymbol{B} = [0 \ldots 0, 1]^T$. The dimension of $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{Q}$ are $n_s \times n_s$, $n_s \times 1$ and $n_s \times 1$ respectively, where n_s is the total number of the states of a cell, which may vary in different ACPs. For example,

$$n_s = (N_1^2 + kN_1 + 2k + 2N_1)/2k, \tag{4.46}$$

for an FCFS ACP. By solving $\boldsymbol{Q} = \boldsymbol{A}^{-1}\boldsymbol{B}$, the state probability $q_{i,j}$ is obtained. Then, all blocking probabilities are obtained numerically by summarizing certain $q_{i,j}$.

The numerical results of the proposed ACPs will be examined in the next section. In general, no matter what type(s) are given priority and to whatever extent, the same analytical method can be used. The system capacity, traffic density and details of an ACP determine the number of states and the transition rates of the Markov chain. The performance of the ACP will then be determined. Furthermore, by selecting a suitable ACP, trade-offs can be made among the blocking probabilities of several types of calls to maximize the amount of traffic handled by a cell under given blocking requirements.

Simulation studies are conducted to simulate the behavior of a cell under some typical ACPs and to verify the analytical results. Simulation takes the following steps: First, select the same parameters used in corresponding analysis, such as N_1, k, k' and m. Second, call arriving streams to a BS are generated as well as the call holding time. All calls are identified by their type. Third, simulate certain ACP. A call will be accepted/blocked depending on ACP rule and l_f at the call arriving instant. Whenever a call is blocked, a blocking event is recorded along with the call's type. In addition, the total number of calls for each type is also recorded. All blocked and dropped calls are eliminated from originally generated call arriving streams such that they have no impacts on the upcoming calls. Finally, blocking probabilities are calculated. Results are reported with corresponding analytical ones.

4. 5 Numerical Results

In this section, some numerical examples are presented to illustrate the performance of the ACPs discussed in previous sections. The following numerical values are used to describe the model of mixed traffic. The total call arriving rates, λ_1 and λ_2, are variables to show ACP performance under different traffic loads. It is assumed that $N_1 = 36$, unless otherwise stated. N_2 changes according to k. We also consider that, in average, all MSs stay in a cell for 1800 s. The average service time is 200 s for both services. Therefore, for each Service 10% of the total traffic is handoff traffic ($\lambda_{h1} \approx 0.1\lambda_1$ and $\lambda_{h2} \approx 0.1\lambda_2$). Different k values are used to show how performances of admission control change as it changes. It is also interesting to see how ACPs perform under different traffic distribution, which is represented by ρ_2/ρ_1.

We number the 36 basic channels in a cell as channel 0 ~ 35. For $k = 2$, we summaries the channel allocation in the ACPs mentioned as in Table 4. 1.

Table 4. 1: Channel allocation under different ACPs

ACP type	FCFS and dropping	Case ($k = 2$)	Reservation ($k = 3$, $m = 1$)
Service 1 new/handoff calls	0 ~ 35	0 ~ 34	0 ~ 32
Service 2 new calls	0 ~ 35	0 ~ 33	0 ~ 29
Service 2 hand-off calls	0 ~ 35	0 ~ 35	0 ~ 35

Nonprioritized ACP

Fig. 4. 1 is the general form of state transition diagram of FCFS ACP, where $kN_2 = 36$ is used to calculate numerical results. The number of states of this two dimensional Markov chain depends on k. An example of all the state probabilities are plotted in Fig. 4. 4 for $k = 3$, $N_1 = 36$ and $N_2 = 12$. The shape of the plot changes according to both service's traffic loads. So as the blocking probabilities. The blocking probabilities of FCFS ACP

with respect to traffic loads and k are shown in Fig. 4. 5. It exhibits that $P_2^{(n)} > P_1^{(n)}$. Simulation results are in good agreement with corresponding analytical results.

As k increases, the fairness factor reduces fast as shown in Fig. 4. 6. This figure is drawn at $P_2^{(n)} = 10^{-2}$ for each k value. To compare the blocking probabilities under different scenarios, the event that the service (or call) type with the highest blocking probability (Service 2 in this case) reaches 1%, is used as a reference. The reasons for doing this are: First, fairness factor is a function of traffic load and therefore blocking probabilities. We should have a common ground for comparison. Second, the designed call blocking rate of a system is in the order of 1%. We are interested in system performance around this blocking level. In several figures shown later, this reference is used again.

Larger k means fewer Service 2 calls can be accommodated in a cell with a fixed N_1 simultaneously. Thus the blocking difference between the two services widens. In Fig. 4. 6, the Service 1 blocking probability varies from more than 40% to less than 5% of that of Service 2, dependent on the value of k. The unequal blocking problem does not exist in a single service system but create new problems in multimedia systems, since service difference may be quite large in those systems. For example, if video service has to face a very high handoff blocking rate than voice service, then few people will be satisfied. In this case, we might have to design alternative ACPs such that video handoff blocking rate is somehow reduced. Although N_2 reduces dramatically as k increases, the actual information carrying capability of a cell does not drop that fast because one Service 2 call equivalent to more than one Service 1 call. However, we do lose multiplex efficiency in Service 2.

We also look at the impacts of both traffic intensity and traffic distribution on the $r_f^{(n)}$ as illustrated in Fig. 4. 7. $r_f^{(n)}$ is not a constant and becomes higher when either $\lambda_1 \neq \lambda_2$ or the traffic load increases. It is because the difference of blocking rate caused by co-existence of two services reduces when one type of traffic dominates. It also suggests that $P_1^{(n)}$, $P_2^{(n)}$, and therefore $r_f^{(n)}$, have different sensitivity to the traffic variations. $P_1^{(n)}$ increases faster than $P_2^{(n)}$ as the total traffic increases. The increase in Service 1 traffic load makes fairness factor increase more. It is suggested to use the condition of $\rho_1 = \rho_2$ (or

$\lambda_1 = \lambda_2$), i. e. equal traffic load, as a reference of the traffic distribution to view the fairness factor variations when $\lambda_1 \neq \lambda_2$. Therefore, we have two references to show performance under typical scenarios and make comparisons.

The major observations on FCFS ACP are: being a simple ACP, it does not protect any calls. This policy might not be suitable when k is too high. Its performance is also affected by traffic distribution. To compare its performance under different scenario, reference is need to be selected.

Prioritized ACPs

Now, let us examine the performances of different prioritized ACPs. First of all, three cases of access limitation on calls of lower degree of importance are given. Equal traffic load condition is considered. Here we assume that $k = 2$ as in Section 4. 4. 1. The results on Service 2 new calls blocking rates of these cases which limit the access of Service 2 new calls and/or Service 1 calls is listed in Table 4. 2.

Table 4. 2: Blocking probability of Service 2 new calls

Traffic (Erl)	2	4	5	6	8
Case 1	3.58×10^{-9}	6.15×10^{-5}	8.27×10^{-4}	5.15×10^{-3}	4.79×10^{-2}
Case 2	3.57×10^{-9}	6.11×10^{-5}	8.18×10^{-4}	5.07×10^{-3}	4.67×10^{-2}
Case 3	3.51×10^{-9}	5.87×10^{-5}	7.77×10^{-4}	4.77×10^{-3}	4.30×10^{-2}

Blocking probability of Service 2 new calls are similar in these three cases. The degree of access limitation is different for Service 1 call but same for Service 2 new calls. The lowest $P_{n2}^{(p1)}$, which is in case 3, is 90% or above of the highest $P_{n2}^{(p1)}$, which is in case 1. However, the relationship between $P_1^{(p1)}$ and $P_{h2}^{(p1)}$ can be changed as shown in Fig. 4. 9, i. e. from $P_1^{(p1)} > P_{h2}^{(p1)}$ to $P_{h2}^{(p1)} > P_1^{(p1)}$. Service 2 handoff blocking probability is reduced but different call types are performed differently. On one hand, the cost on Service 2 new call blocking is similar. On the other hand, $P_1^{(p1)}$ is increasing as the case number increases due to its access limitation increasing. As we gradually limit Service 1, Service 2 handoff gains. This gain is only marginal between Case 1 and Case 2, but quite significant between Case 2 and Case 3. The reason is that, in Case 3, when $l_f = 2$ only Service 2 handoff calls can be assigned a channel (2 basic channels). The competition with Service 2 handoff call from other calls is avoided. Although the example is shown for $k = 2$, it can be extended to other k values. These cases show that we can rather flexibly change the blocking probability of each type of calls. By carefully closing certain access limitations for certain type of calls, important calls can be protected.

Channel reservation was originally proposed for single traffic systems and is reexamined here in a heterogeneous traffic scenario. Basically, reservation is still a kind of putting certain access limitation onto certain type of calls. But it is more narrow concept compared to the access limitation. For example, the discussed Case 1 and Case 2 are not reservation schemes. Fig. 4. 2. (a) gives the channel allocation among three coming traffic types. Fig. 4. 2 (b) shows the states and state transition rates of the Markov chain of an ACP which reserves mk basic channels for Service 2 handoffs. The cell states in the chain, which are the same (same means having the same number of neighbors and the transition rate from or to their neighbors being correspondingly labeled the same. Of course, the value of $q_{i,j}$ in FCFS ACP and that in this ACP are different) compared to their counterpart states in the FCFS ACP chain, are not shown for simplicity. We would like to mention that the top, down, left and right neighbors of a state (i,j) are states $(i,j-1)$, $(i,j+1)$, $(i-1,j)$, and $(i,j+1)$, respectively. Of course, not all of the shown states have four neighbors.

The performance of this ACP is compared with that of the FCFS ACP, a special case of channel reservation with $m = 0$. In Fig. 4. 10, the significant costs on the rest types of calls, i. e. Service 2 new calls and Service 1 calls, are shown compared with the FCFS ACP case. $P_1^{(n)}$ and $P_2^{(n)}$ are about 50% or lower of $P_{n1}^{(p2)}$ and $P_{n2}^{(p2)}$, respectively. The performance improvement on Service 2 handoff calls is in Fig. 4. 11. It is found that $P_{n2}^{(p2)}$ is still the highest because Service 2 new calls are not protected and unable to access all basic channels. As k increase, the amount of traffic carried will reduce under a given $P_{n2}^{(p2)}$. The simulation results show a very low Service 2 handoff blocking rate. For 200,000 calls simulated ($m = 1$), only 1 Service 2 handoff blocking is observed when $\rho_1 = \rho_2 = 6$ and 4.5 Erlangs, but no such blocking is observed under 3.6 Erlangs traffic.

$r_{f1}^{(p2)}$ reduces as k increases (in a similar way as $r_f^{(n)}$ in Fig. 4. 6). The $r_{f1}^{(p2)}$ is higher compared to $r_f^{(n)}$ as in Fig. 4. 12, indicating that fairness is improved between Service 2 new calls and all Service 1 calls. The $r_{f2}^{(p2)}$ approaches 0 very quickly as m increases. That means if we reserve sufficient number of basic channels, Service 2 handoff blocking will become negligible compared to other call blockings. Therefore, to select an m, trade-off should be made between the $P_{h2}^{(p2)}$ and other blocking probabilities.

It is found that $P_{h2}^{(p2)}$ is very sensitive to the variations of the traffic distribution, $\rho_2/\rho_1 = \lambda_2/\lambda_1$. In order to facilitate a fair comparison among different traffic distributions, the reference $P_{n2}^{(p2)} = 10^{-2}$ is used. As shown in Fig. 4. 13, the highest value of $P_{h2}^{(p2)}$ is higher than its lowest value for about $10 \sim 10^2$ times. As λ_2/λ_1 increases, more and more Service 2 handoffs arrive. Thus under a fixed m, $P_{h2}^{(p2)}$ becomes higher. This observation suggests a possible adaptive channel reservation plan for networks with fluctuate traffic distribution. On one hand, by reducing m when Service 1 traffic dominants, Service 1 blocking is reduced. On the other hand, by increasing m, we can keep $P_{h2}^{(p2)}$ low enough when Service 2 traffic dominants. The increase of $P_1^{(p2)}$ as λ_2/λ_1 increases is also observed. To use the channel reservation ACP, we should bear in mind the incurred cost, especially when k is large.

Finally, Fig. 4. 14 exhibits that, by reserving channels for the entire Service 2, the blocking relationship changes from $P_2^{(p2)} > P_1^{(p2)}$ to $P_2^{(p2)} < P_1^{(p2)}$. $P_1^{(p2)}$ experiences a very significant increase. Therefore, the cost on Service 1 is very high, especially when k and/or m are larger. Since the traffic distribution has a very significant impact on the ACP performance, the selection of m should be based on traffic distribution. The higher the Service 2 traffic load percentage, the higher the m. By doing so, $r_{f1}^{(p2)} = P_1^{(p2)}/P_2^{(p2)}$ can be adjusted. $|P_1^{(p2)} - P_2^{(p2)}|$ can be minimized if necessary.

The ACP with droppings is a new approach of admission prioritization. Any call can access any the basic channels regardless its type. But a BS may drop a Service 1 call to make room for coming Service 2 handoff. A Fig. 4. 15. shows one of its cases. It is found, as expected, that $P_{h2}^{(p3)}$ is reduced significantly while the ongoing call dropping rate, i. e. the cost on Service 1 ($P_{h1}^{(p3)} - P_{n1}^{(p3)}$), is very low. We also compare this ACP and FCFS ACP under same traffic load. It is found that $P_{n2}^{(p3)}$ is no higher than 102% of $P_2^{(n)}$ and $P_{n1}^{(p3)}$ is no higher than 104% of $P_1^{(n)}$. Therefore the cost of call droppings is very low. In a single traffic system, there is no gain but loss to drop a call for a same type call. So this approach is only for systems carrying multiple services.

The selection of k' and the performance improvement of $P_{h2}^{(p3)}$ is shown in Fig. 4. 16. Simulation results are well fit with the analytical results. Similarly, we still set $P_{n2}^{(p3)}$ to be 0.01 as a condition to compare blocking probability under different k'. The improvement on $P_{h2}^{(p3)}$ slows down as the k' increases while the cost on Service 1 just increases slightly. In Fig. 4. 17, the average number of call droppings is shown to be a function of traffic loads. This figure reveals why Service 1 dropping is not high. The probability of dropping one call is (much) higher than those of dropping multiple calls at a time. When traffic increases, $P_{d,l}^{(p3)}$ with a higher l increases its percentage so $n_{d1}^{(p3)}$ increases as well. However, $n_{d1}^{(p3)}$ is not sensitive to traffic variation. Moreover, while k' varies from 1 to 5, $n_{d1}^{(p3)}$ has a range from 1 to 1.85, which gives a possible range for the θ_c selection.

The impact of the traffic distribution on $r_{f3}^{(p3)}$ is illustrated in Fig. 4. 18. To select a higher k' loses the fairness further within the Service 2 as more handoff calls are saved. Under different traffic distribution, $r_{f3}^{(p3)}$ reacts differently. When Service 1 traffic dominates, fewer Service 2 handoffs are generated due to a lower Service 2 traffic load in the system. The opportunity of dropping Service 1 calls is thus not high. When Service 2 traffic becomes dominate, there might not be enough Service 1 calls to be dropped so $r_{f3}^{(p3)}$ increases for all k', indicating the advantage of dropping (offer better handoff protection) diminishes. However, significant $r_{f3}^{(p3)}$ reduction can be seen for certain λ_2/λ_1 and k' values. For the case of $k' = k-1 = 3$, $r_{f3}^{(p3)}$ is low only when $\lambda_2/\lambda_1 < 3$. As Service 1 traffic load becomes too low, dropping three Service 1 calls simultaneously is rare. Performance of $k' = 3$ is approaches performance of $k' = 2$ as λ_2/λ_1 approaches 10.

Given $P_{n2}^{(p3)} = 10^{-2}$, we draw the traffic amount of each service under different λ_2/λ_1 in Fig. 4. 19. As expected, the traffic amount of Service 2 is less sensitive to λ_2/λ_1

variations under a given $P_{n2}^{(p3)}$. In other words, with the traffic distribution changing, traffic of Service 2 experiences less changes compared to that of Service 1.

The advantage of this ACP is that a significant improvement in Service 2 is at little cost. When k is higher, ACP with droppings will gain more since the average number of calls dropped each time is relatively smaller compared to k. The ACP with droppings has a limitation. It can not make Service 2 handoff blocking rate lower than that of the Service 1. If this is the design requirement, channel reservation has to be used.

Finally, we provide an example of using soft capacity to maximize fair access between two services, which is based on the DS/CDMA system discussed in Chapter 3, is also given. It is found that $P_1^{(n)} < P_1^{(p4)} = P_2^{(p4)} < P_2^{(n)}$ as illustrated in Fig. 4. 21. The probability that transmission quality is below the QoS due to soft capacity is quite low (as $P_1^{(p4)} = P_2^{(p4)} = 10^{-2}$, $P_{sc}^{(p4)} \approx 2.5 \times 10^{-3}$). In another example, the blocking rate of the same DS/CDMA system with different data power allocation, m_2, is shown in Fig. 4. 22. As m_2 reduces, the difference between $P_1^{(p4)}$ and $P_2^{(p4)}$ shrinks. Since the cell capacity reduces as m_2 increases, blocking of both service calls also increases under a given traffic load. Here we use the same analytical method (i. e. Markov chains) to find above results under arbitrary k values ($1 < k < 2$). This example directly relates power allocation and call blocking probabilities. From the example, we see that the unique CDMA feature, soft capacity, can be used as an alternative of ACP design. The probability that the system runs into soft capacity is reasonably low. Power level, as well as power control error, in CDMA systems decides the resource requirement to support of a service. When considering traffic handling capability of a CDMA system, we must know the power control performance.

4. 6 Conclusions

As multi-service is going to be carried by the next generation wireless cellular networks, admission control under mixed traffic scenario becomes one of the important issues. In this chapter, two dimensional Markov chain models are used to analyze the blocking probability of both services. Results show that the blocking probability of the services are generally different. By giving certain type(s) priority, we can improve their service quality by reducing their blocking rate. Handoff calls of the high blocking rate service are chosen to give the priority over the rest.

The major conclusions of this chapter are:

- Different services suffer different blocking probability under FCFS ACP. As the difference increases, blocking differences also increases. FCFS ACP is simple, but performance of certain type of calls may not be satisfactory.
- Prioritization in ACPs is necessary due to the different degree of importance among multiple services. Either channel access limitation or enhancement can be used.

  ACP with dropping ongoing calls reduces handoff failure of high blocking rate service only to a limited degree but at very little cost. Multiple call drops are rare. This ACP is proposed only for multiple service systems.

  ACP with channel reservation can greatly reduce handoff blocking rate at a rather high cost. The number of reserving channels should not be too high and may be adaptive.

  The soft capacity feature of DS/CDMA can be utilized to provide fair access.
- Traffic distribution will affect ACP performance, in terms of both blocking probability and fairness in access. The design and selection of an ACP should consider this factor. Traffic models for the future systems should be investigated.

These mentioned ACPs can achieve different design goals based on system blocking requirements. They are rather flexible. We can extend prioritization from certain handoff to more types of calls as well. However, when the performances of each type of calls are dependent, it is evident that trade-offs is necessary for the ACP design. We have to be careful on setting policies to meet our requirements.

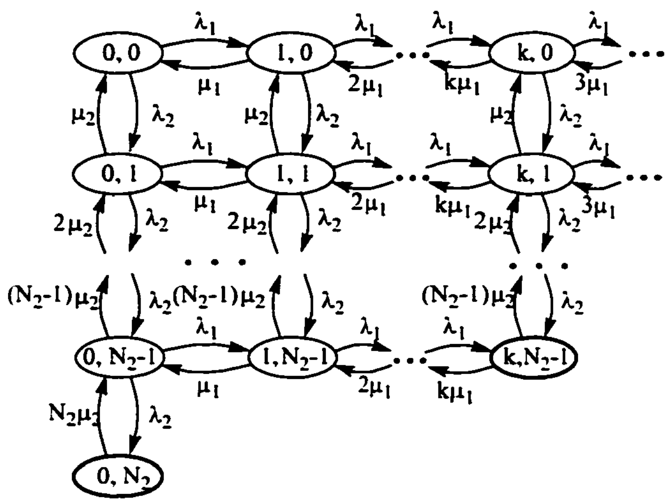
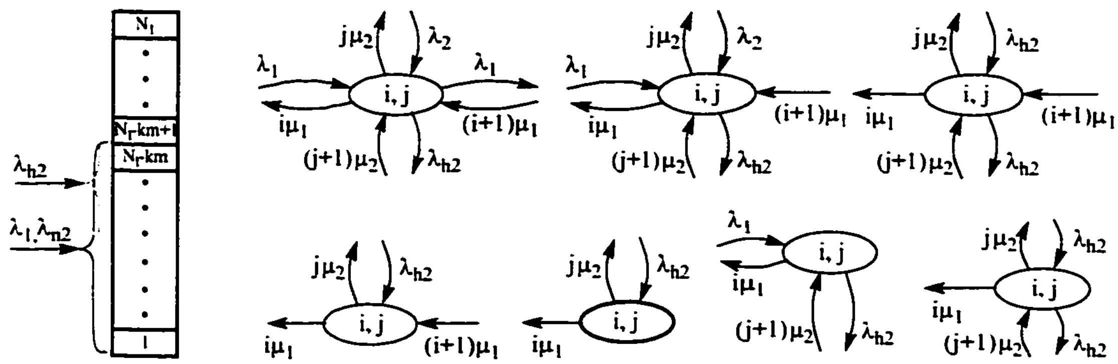
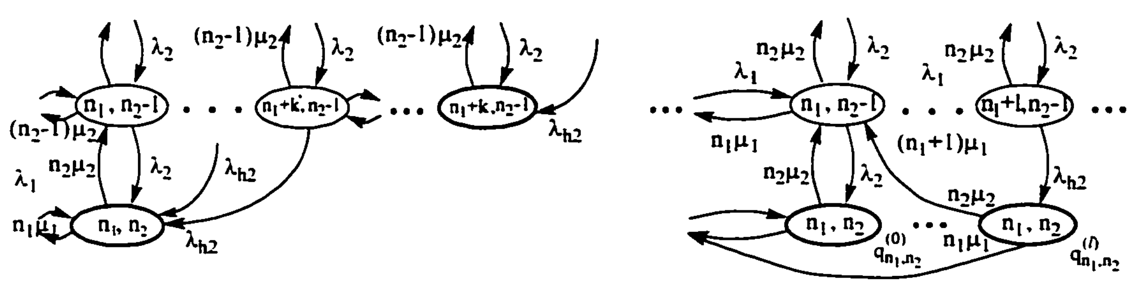


Fig. 4. 1 Markov chain model for the first-come-first-served ACP.



(a) (b)

Fig. 4. 2 Markov chain model for the ACP with channel reservation for Service 2 handoff traffic.



(a) (b)

Fig. 4. 3 Markov chain model for the ACP with Service 1 call dropping for Service 2 handoff calls.

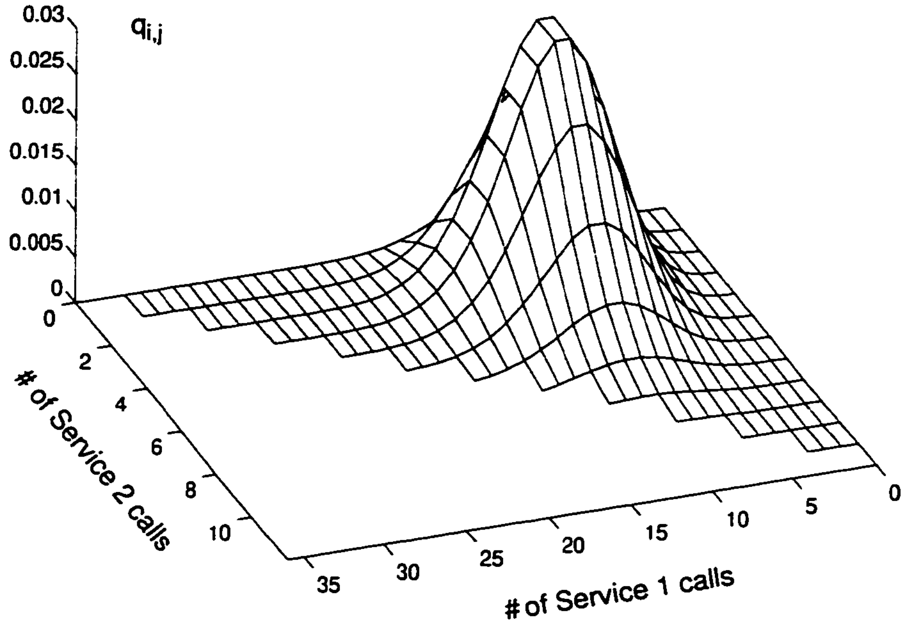


Fig. 4. 4 State probabilities of FCFS ACP (k = 3). 10 and 3 Erlangs of traffic for Service 1 and 2 is in a cell respectively.

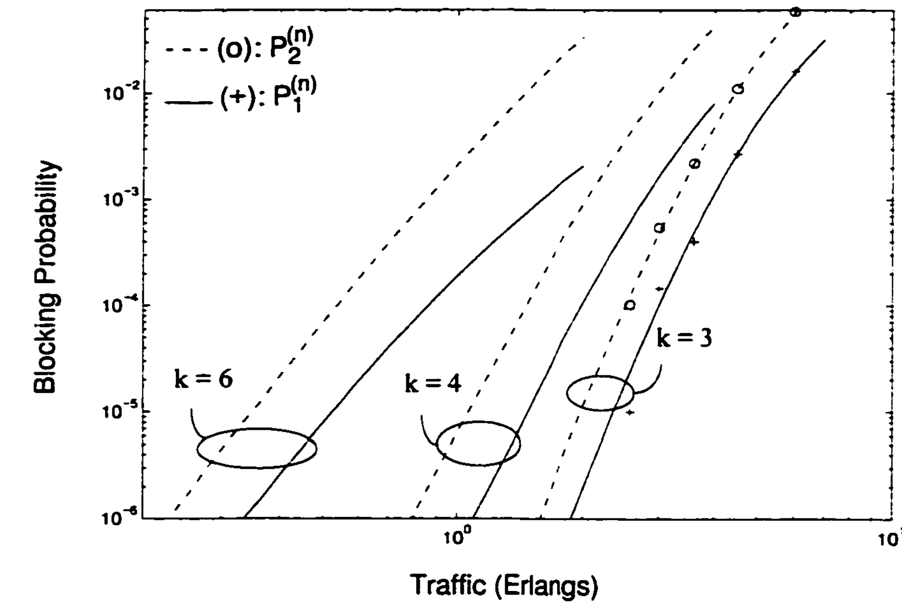


Fig. 4. 5 Blocking probabilities of Service 1 and Service 2 under FCFS ACP ($\rho_1 = \rho_2$). Dots are simulation results.

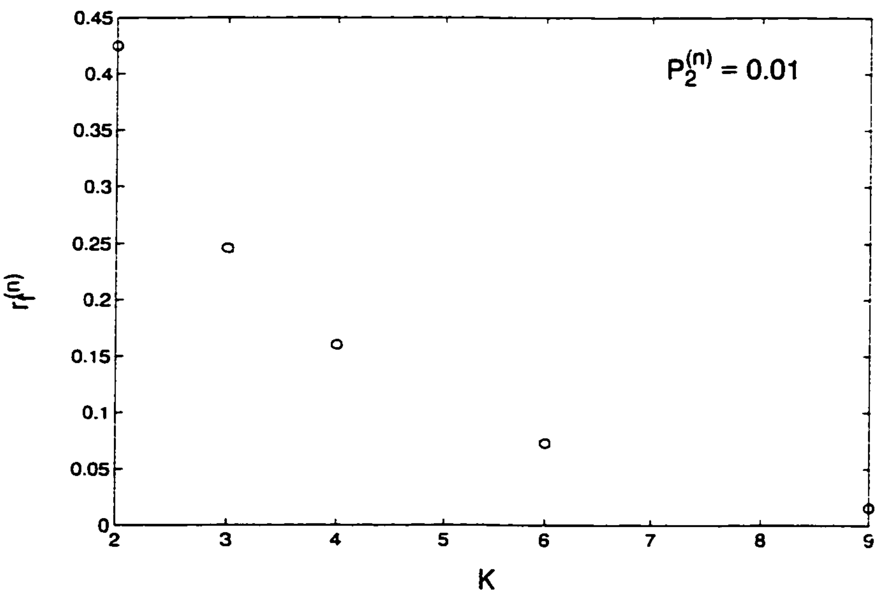


Fig. 4. 6 Fairness factor of FCFS ACP changes under different k ($\rho_1 = \rho_2$).

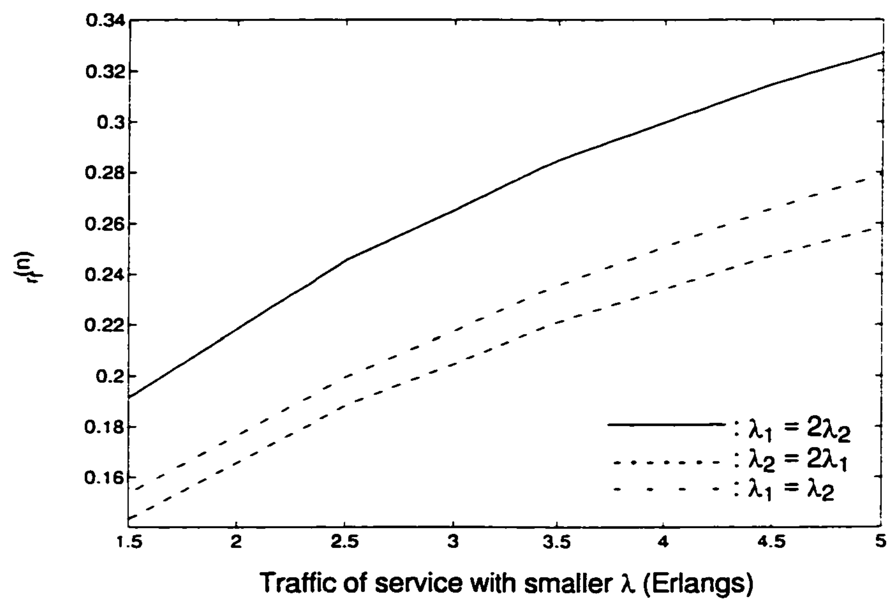


Fig. 4. 7 The fairness factor of FCFS ACP under different traffic distribution ($k = 3$).

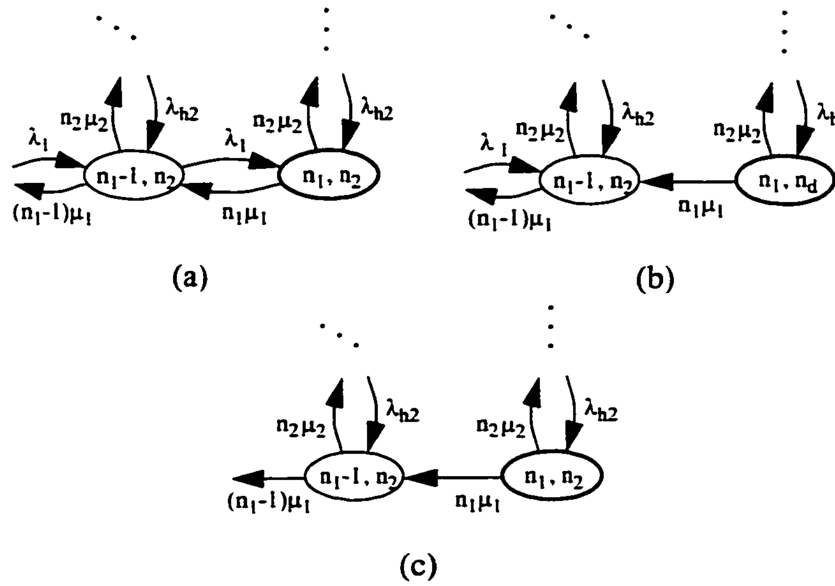


Fig. 4. 8 The Markov chains of Case 1 (a), 2 (b), and 3 (c). Only the part of the chain, which is different from the corresponding part in Fig. 4. 1 due to the ACP adjustment, is shown.

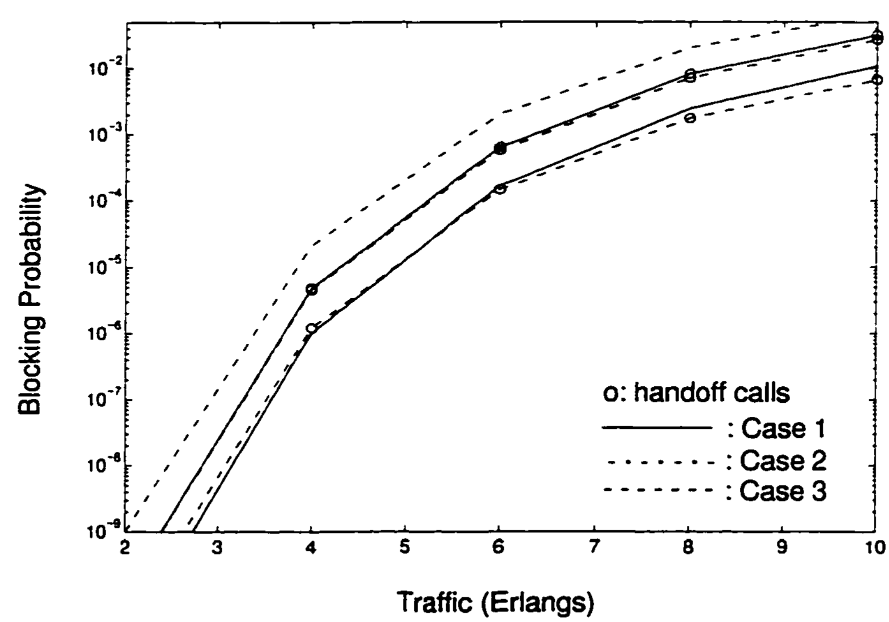


Fig. 4. 9 Comparison of handoff schemes (i.e. Case1, 2 and 3). $\lambda_1 = \lambda_2$.

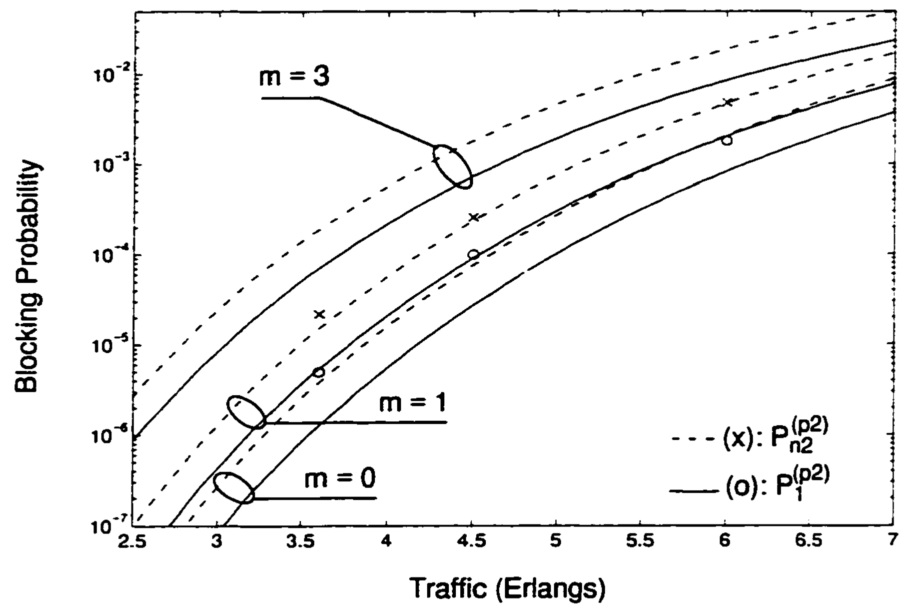


Fig. 4. 10 Blocking probabilities under ACP with/without channel reservation ($k = 2$, $\rho_1 = \rho_2$). Dots are simulation results for $m = 1$.

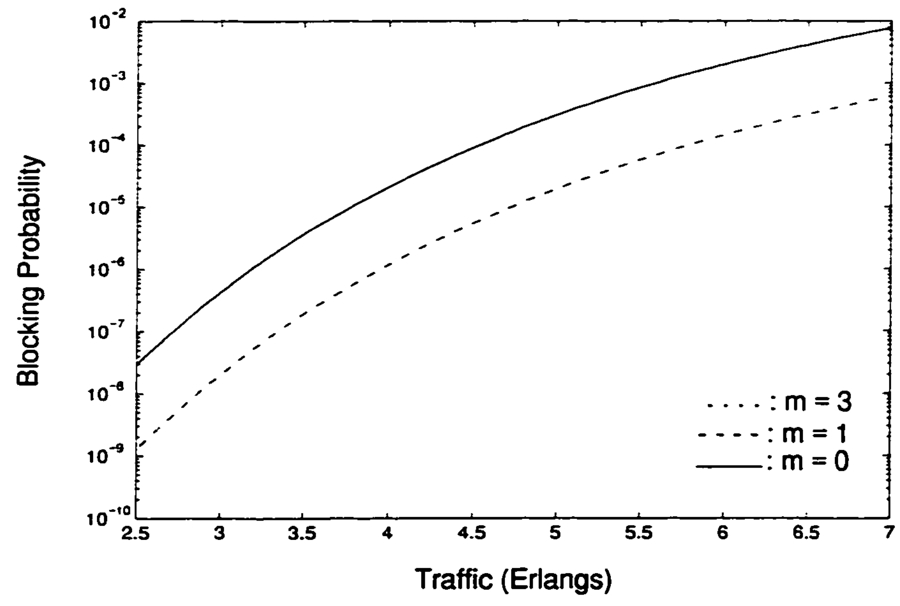


Fig. 4. 11 Blocking probabilities of Service 2 handoff under ACP with/ without channel reservation ($k = 2$, $\rho_1 = \rho_2$).

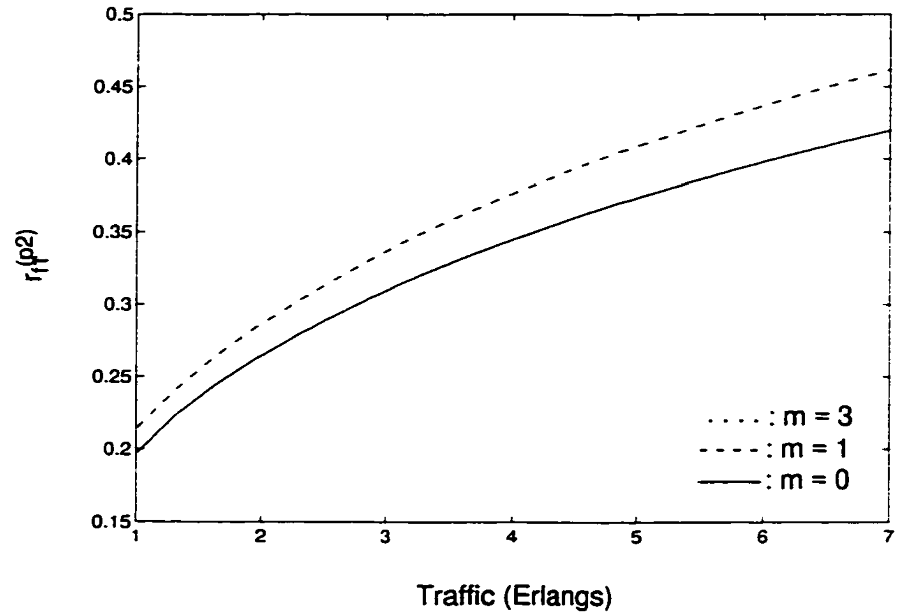


Fig. 4. 12 Fairness factor of ACP with/without channel reservation ($k = 2$, $\rho_1 = \rho_2$).

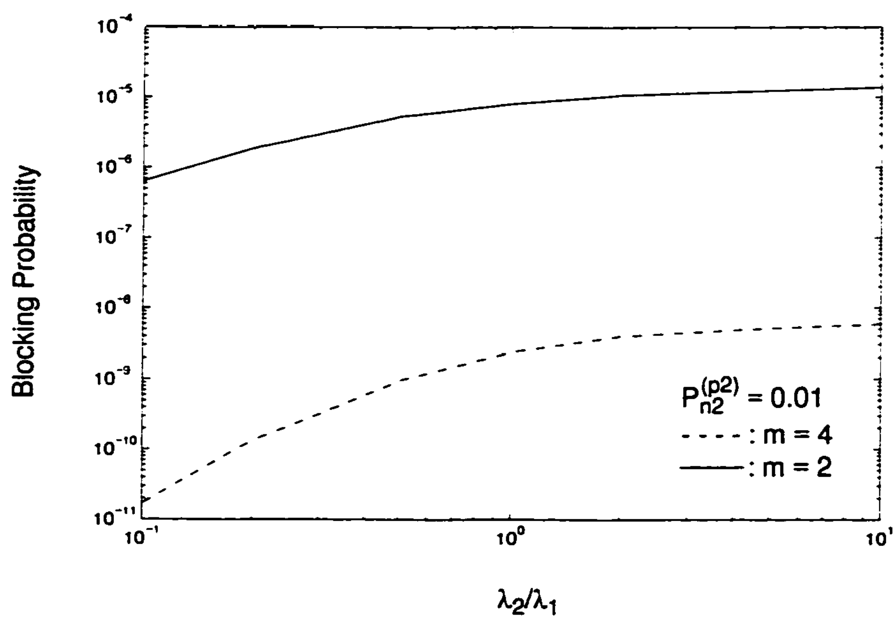


Fig. 4. 13 Service 2 handoff blocking rate of ACP with channel reservation ($k = 3$).

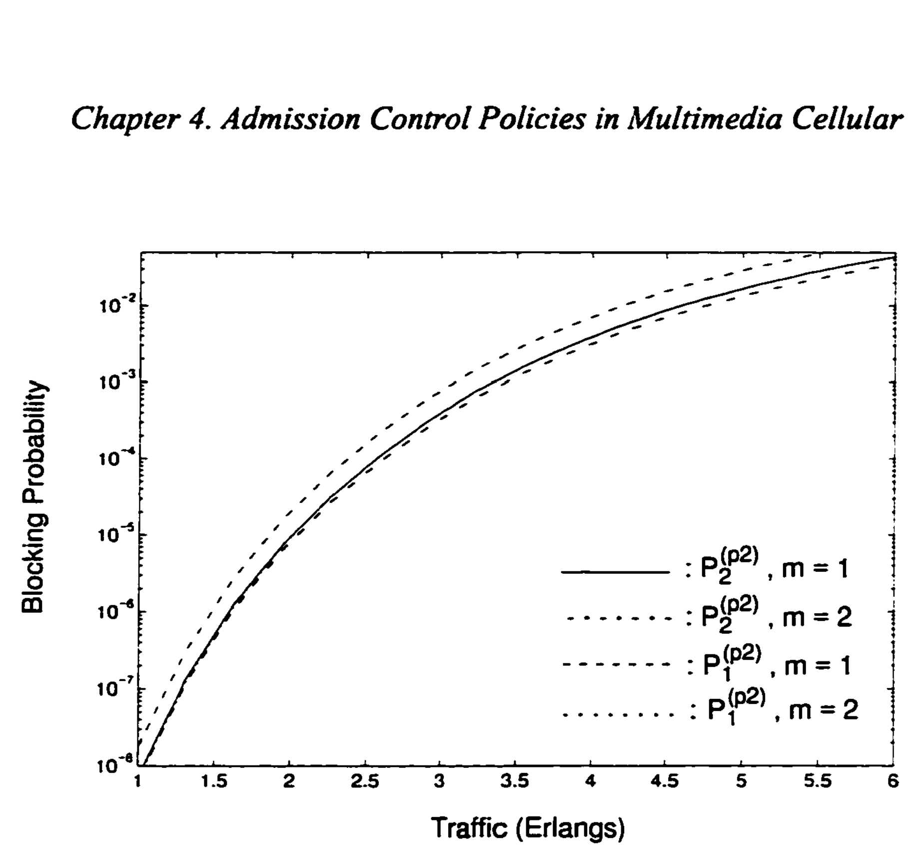


Fig. 4. 14 Blocking probability of ACP with channel reservation for Service 2 ($k = 3$, $\rho_1 = \rho_2$).

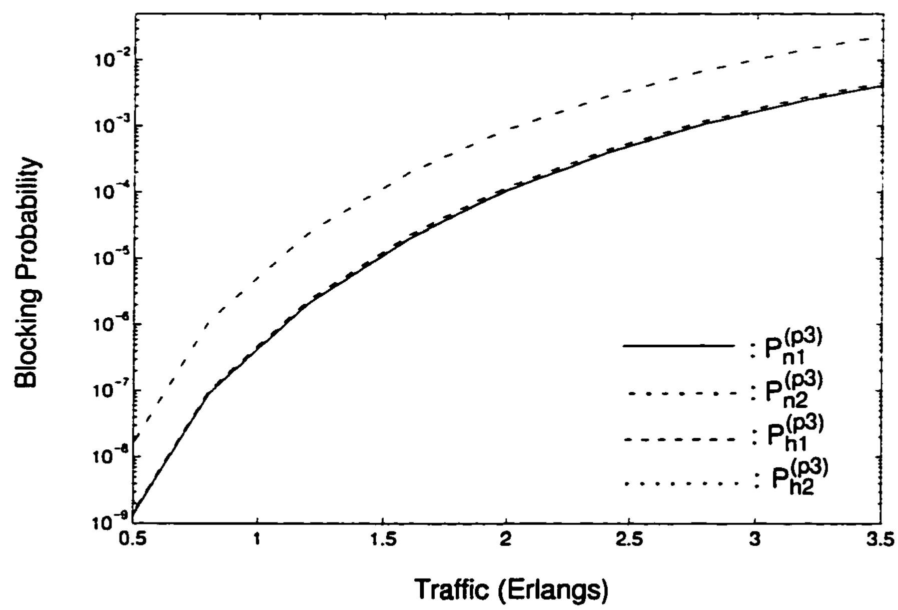


Fig. 4. 15 Blocking probabilities of the ACP with droppings under different traffic amount ($k = 4$, $k' = 2$, $\rho_1 = \rho_2$).

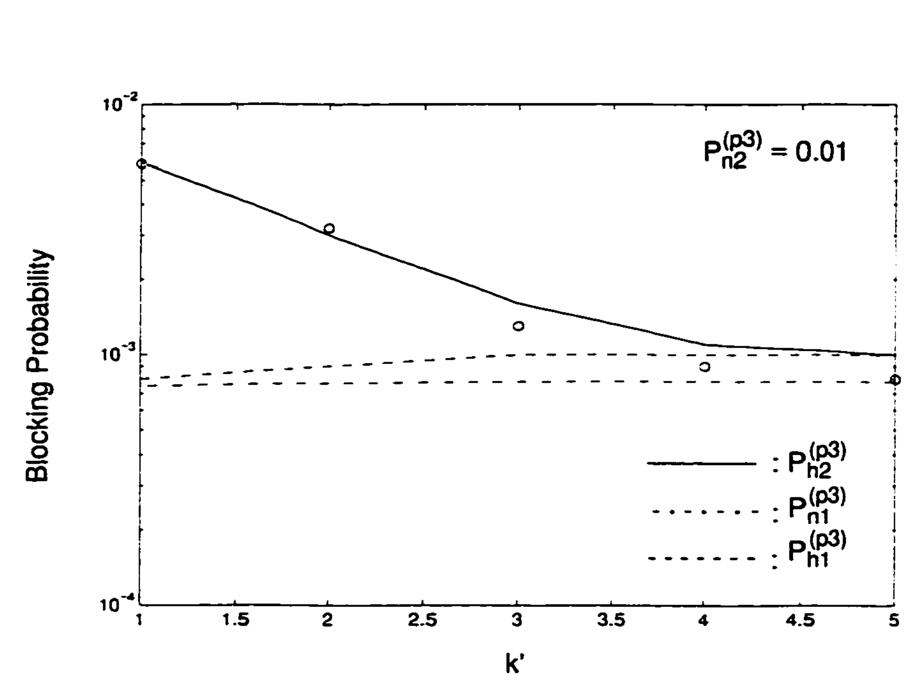


Fig. 4. 16 Blocking probabilities of ACP with Service 1 call droppings ($k = 6$, $\rho_1 = \rho_2$). Simulation results are marked by 'o' for Service 2 handoffs.

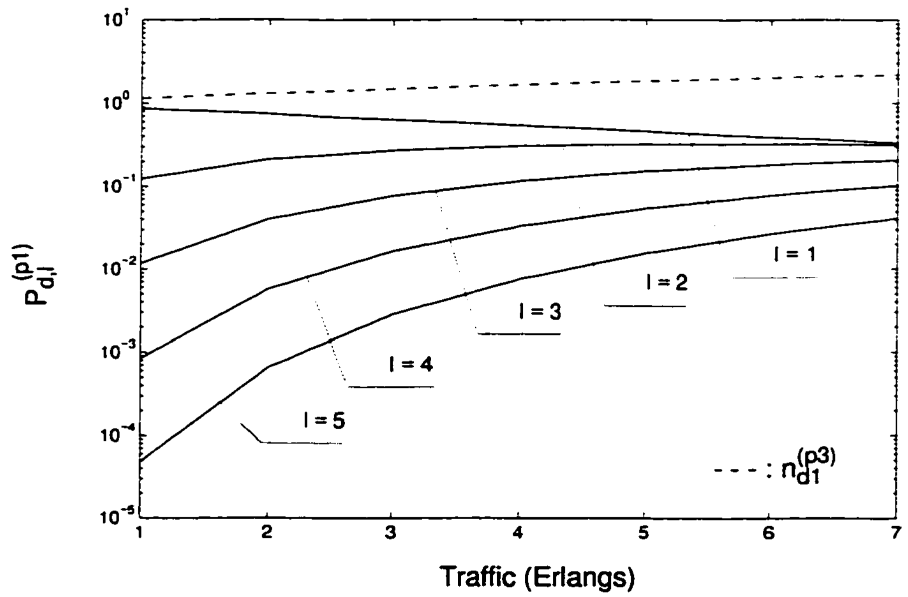


Fig. 4. 17 Average number of droppings and the probability of dropping l calls ($k = 6$, $k' = 5$, $\rho_1 = \rho_2$).

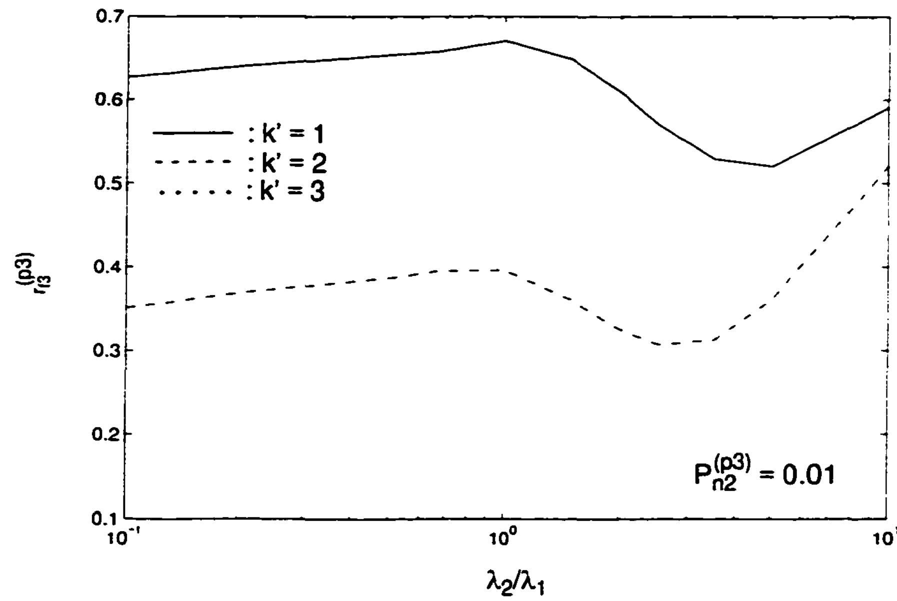


Fig. 4. 18 Fairness factor of ACP with droppings under different k' ($k = 4$).

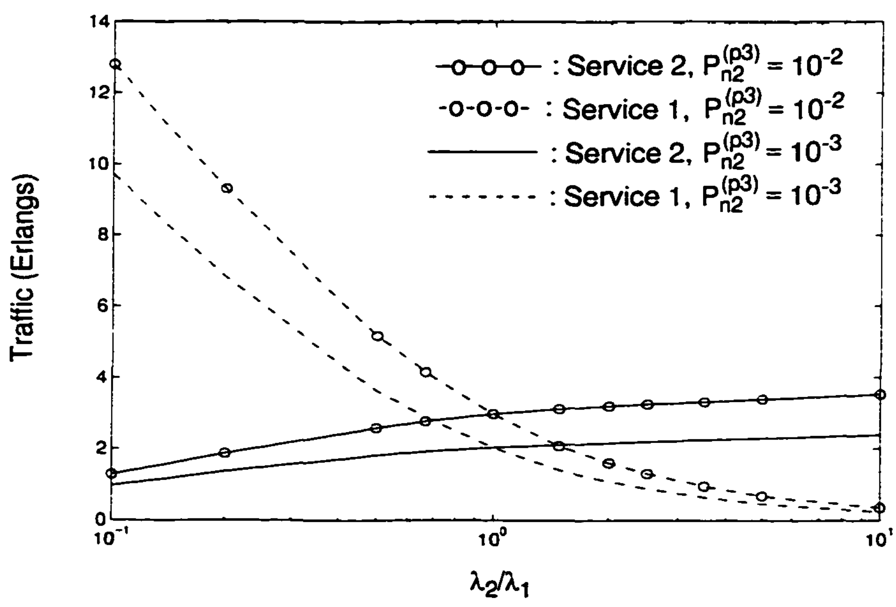


Fig. 4. 19 Traffic of both services under ACP with droppings ($k = 4$, $k' = 2$).

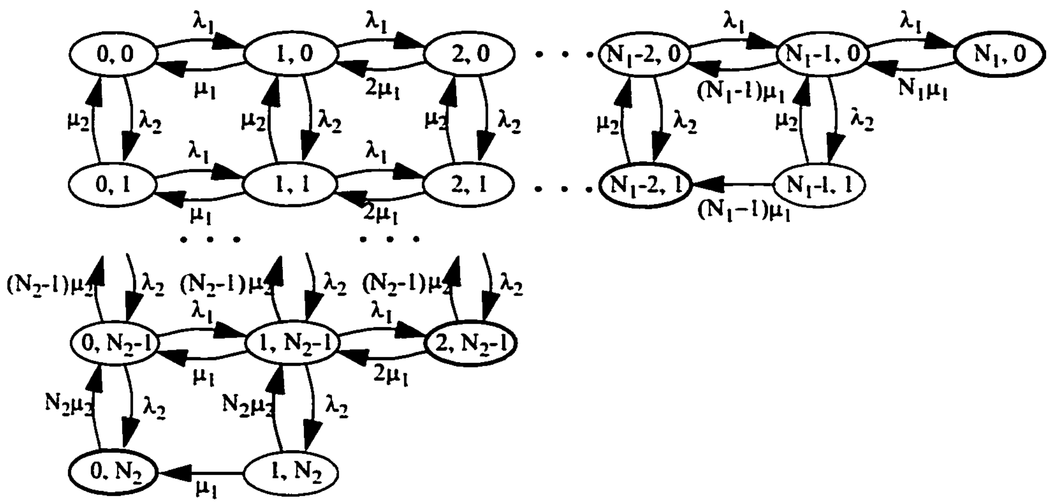


Fig. 4. 20 Markov chain model of prioritized ACP using soft capacity ($k = 2$).

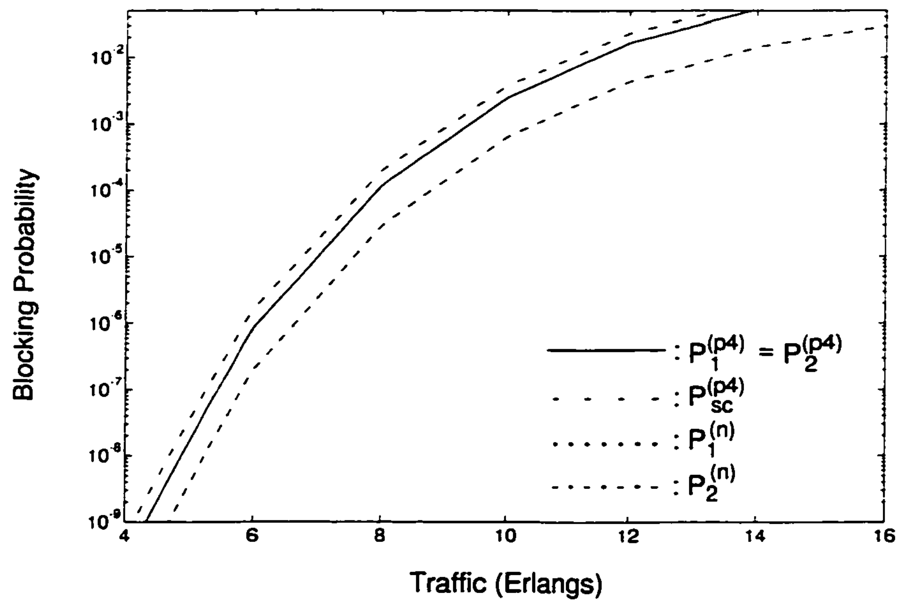


Fig. 4. 21 To ensure fairness between services by soft capacity. Blocking probabilities of voice service (Service 1) and data service in a DS/CDMA cellular system ($\lambda_1 = \lambda_2$). $\alpha_1 = 3/8$ and $\alpha_2 = 0.45$. Other parameters are the same as in Chapter 3 Case 1.

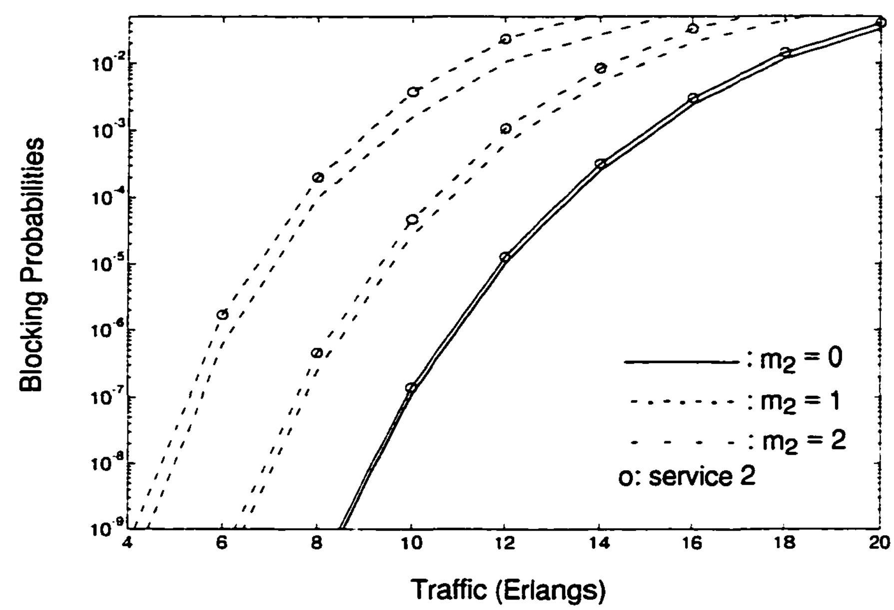


Fig. 4. 22 Blocking probabilities of voice service (Service 1) and data service in a DS/CDMA cellular system ($\lambda_1 = \lambda_2$). FCFS ACP. $\alpha_1 = 3/8$ and $\alpha_2 = 0.45$. Other parameters are the same as in Chapter 3 Case 1.

Chapter 5

Handoff Issues in Cellular Systems

5. 1 Handoff Strategies in a Hierarchical Cellular System

Since the entire service area of a cellular system is partitioned into many cells, handoffs are unavoidable. The methods of serving handoff better have been studied for systems with cells of the same size [9], [17]. For future systems, services will be provided to users of high mobilities.

The capacity demand requires the deployment of microcells for a high degree of frequency reuse. However, fast moving MSs may generate a lot of handoffs if the cell size is small. The increase in handoff signalling load and the call dropping rate is burdensome. So microcellular communication systems with hierarchical macrocell overlays are proposed. Cells of different sizes coexist in one system. Macrocells can also provide overflow channels for microcells and/or cover low traffic areas. Several authors have discussed the system blocking performance of using overlaid macrocells as overflow channel providers, the evaluation of channel allocations criteria, and the approach to share the spectrum between two tiers of cells to maximize the system capacity [3], [7], [13]. A handoff algorithm minimizing the handoff rate in a multi-tier cell layout is proposed without giving the performance gain figures [10]. It is clear that high-speed MSs have priority in accessing macrocells is the only way to reduce handoff rate further. Our aim is to study and compare the impact of user mobilities on the system performance by analyzing two different traffic management strategies.

5. 1. 1 Analytical model

System model

A hierarchical system is composed of microcells and macrocells with radius R_1 and R_2, respectively. For each macrocell there are seven microcells underneath as in Fig. 5. 1. So $R_2 = \sqrt{7}R_1$. Assume that there are C_1 and C_2 channels in microcell and macrocell, respectively, and this can give a certain blocking probability during peak hours. No guard channel is given to handoffs. Blocked new calls and handoff calls are cleared.

It is also assumed that a large number of MSs travel in the region covered by these two tiers of cells. Suppose the moving speed v of an MS is constant during a call, and a random variable uniformly distributed with pdf:

$$f_V(v) = \begin{cases} 1/V_{max} & \text{for } 0 \le v \le V_{max} \\ 0 & \text{else} \end{cases}. \tag{5.1}$$

Two strategies are examined to see the performance gain of using micro/macrocell to serve low/high-speed users. Strategy 1, based on a threshold speed V_T, classifies MSs into two groups, $[0, V_T]$ and $[V_T, V_{max}]$. MSs in low/high-speed group are assigned microcell/ macrocell channels. When an MS initializes a new call or a handoff call, the BS selects a tier to serve the MS according to its speed. A call will be blocked (or forced to terminate during handoff), if no free channel is available in the tier selected for that call. An occupied channel will be released upon a call is completed or an MS moves out of the cell. No overflow between tiers is considered. The value of V_T guarantees balanced traffic between tiers (same call blocking probabilities in the two tiers).

In strategy 2, a high-speed MS may use a microcell channel and a low-speed MS may use a macrocell channel. In other words, all MSs are treated equally regardless of their speed, but the traffic balance between the tiers is maintained by the following method. The probability of choosing a free macrocell/microcell channel is equal to the probability of a MS being in the high/low-speed group. So in strategy 2 the probabilities of MS being served by each tier are equal to their counterparts in strategy 1. In other words, the traffic loading level of each tier is the same for both strategies. So we can make a fair comparison of their performance.

Traffic model

The call holding time is the amount of time that the call would remain in progress if it could continue to completion without handoff failure. We let the call holding time have the probability density function $f_c(t) = \mu e^{-\mu t}$ for $t \geq 0$, where $1/\mu$ is its mean.

By using the traffic model proposed in [10], the average handoff occasions (average number of handoffs) of a call $E[n] = \bar{n}$ is a function of MS's speed v and cell size R:

$$f_{\bar{n}}(v, R) = \frac{\pi v}{4\mu R} + \frac{\pi v}{3\pi v + 8\mu R}. \tag{5.2}$$

The second term is very small compared with the first term and is neglected in the following analysis. The dwell time is the amount of time an MS remains within a cell. It has an exponential distribution with mean of $1/\eta$. Of course the η is a function of MS's speed v and cell size R $(R = R_1$ *or* $R_2)$. The average dwell time $1/\eta$ is $1/\mu$ divided by the average handoff rate:

$$\eta \approx \pi v/4R. \tag{5.3}$$

Both new call and handoff call arrival processes are assumed to be Poisson process. The total new call arriving rate in a microcell area is λ. Let λ_i be the new call arriving rates in a cell, where the subscript $i = 1, 2$ indicates microcell and macrocell parameters respectively throughout this section. From Fig. 5. 1, we have $\lambda = \lambda_1 + \lambda_2/7$ where:

$$\lambda_1 = \lambda \int_0^{V_T} f_V(v)\, dv \text{ and } \lambda_2 = 7\lambda \int_{V_T}^{V_{max}} f_V(v)\, dv\ . \tag{5.4}$$

Let p_{bi} be the blocking probability of the new call attempts and $E[n_i]$ be the average number of handoffs that a call experienced before termination for each tier. We have the total handoff rate λ_h and the handoff rate for each tier as in [17]:

$$\lambda_h = \lambda_{h1} + \lambda_{h2} = (1 - p_{b1}) E(n_1) \lambda_1 + (1 - p_{b2}) E(n_2) \lambda_2 . \tag{5.5}$$

The total traffic carried within a microcell area is $\rho = \lambda / \mu$ Erlangs. The number of MSs in a cell is assumed to be much larger than the number of the channels in that cell such that the call arriving rate is constant. Since the system handles a handoff call exactly the same as a new call, the forced termination probability p_{fi} is equal to p_{bi}:

$$p_{fi} = p_{bi} = \frac{(\lambda_i + \lambda_{hi})^{C_i}}{(\mu + \eta)^{C_i} C_i!} \Big/ \sum_{k=0}^{C_i} \frac{(\lambda_i + \lambda_{hi})^k}{(\mu + \eta)^k k!} . \tag{5.6}$$

In realistic situations, both p_{bi} and p_{fi} are much smaller than 1 resulting in, from Eq. (5.5), $\lambda_{hi} \approx E(n_i) \lambda_i$. Therefore, $(\lambda_i + \lambda_{hi}) / (\mu + \eta) \approx \lambda_i / \mu$, which means that both p_{bi} and p_{fi} can be viewed as independent of user speed and cell size.

5. 1. 2 Performance analysis

Handoff rate reduction

Let the probabilities of an MS entering the microcell tier or macrocell tier be P and $1 - P$ respectively. For strategy 1, the average number of handoffs of a call, whether it is completed or not, is denoted by $\Re_1$ and is given by [8]:

$$\Re_1 = P \int_0^{V_T} f_{\bar{n}}(v, R_1) f_L(v) dv + (1 - P) \int_{V_T}^{V_{max}} f_{\bar{n}}(v, R_2) f_H(v) dv . \tag{5.7}$$

In Eq. (5.7), we denote $f_L(v) = f_V(v | v \le V_T)$ and $f_H(v) = f_V(v | V_T < v \le V_{max})$. The two terms in Eq. (5.7) represent the contribution of low/high-speed MSs.

In strategy 2, the situation is more complicated as an MS may be served by either tier. So both the high speed group and the low speed group are further divided into two subgroups. Each subgroup is a part of group traffic in a certain tier.The average number of handoffs of a call in strategy 2, $\Re_2$, can be derived as:

$$\begin{aligned}\Re_2 = {} & P^2\int_0^{V_T} f_{\bar{n}}(v, R_1) f_L(v)\,dv + P(1-P)\int_0^{V_T} f_{\bar{n}}(v, R_2) f_L(v)\,dv + \\ & P(1-P)\int_{V_T}^{V_{max}} f_{\bar{n}}(v, R_1) f_H(v)\,dv + (1-P)^2\int_{V_T}^{V_{max}} f_{\bar{n}}(v, R_2) f_H(v)\,dv\end{aligned} \qquad (5.8)$$

The four terms in Eq. (5.8) represent contribution of the four subgroups. The difference between the average number of handoffs is $\Delta\Re = \Re_2 - \Re_1$. Fig. 5. 3 (a) shows the relationship of average number of handoffs and V_T/V_{max} under different V_{max}.

Since V_T directly controls the partition of the total traffic into different tiers, an inappropriate V_T may results in overloading one tier and underloading the other. As all MSs have the same priority regardless of the tier they reside, the blocking probability p_b and the forced termination probability p_f at each tier should be equal. Let $p_b = p_{b1} = p_{b2}$ and $p_f = p_{f1} = p_{f2}$. We first get ρ_1 and ρ_2, the maximum traffic carried by a microcell and a macrocell, respectively. Then, from the system configuration, V_T is given by:

$$V_T = \frac{7\rho_1}{\rho_2 + 7\rho_1} V_{max}. \qquad (5.9)$$

The system capacity is defined as the amount of total traffic, excluding handoff traffic, carried by both tiers under a given p_b, within a certain area (a microcell in this work). For both strategies under the same p_b restriction, these capacities are nearly equal, but their probabilities of call not completed are quite different.

Probability of call not complete

When MS dwell time is exponentially distributed, the probability of a call attempt not completed by either blocking or handoff failure is [17]:

$$p_{nci} = 1 - \frac{1 - p_{bi}}{1 + \eta p_{fi}/\mu}. \qquad (5.10)$$

From Eq. (5.3) and (5.10), we have:

$$p_{nci} = \frac{4R_i\mu p_{bi} + \pi v p_{fi}}{4R_i\mu + \pi v p_{fi}}. \tag{5.11}$$

The smaller the cell size and the higher the MS's speed, the higher the p_{nci} and its mean $\bar{p}_{nci}$ averaged over [0, V_{max}]. To compare the performance of the two strategies, we first calculate $\bar{p}_{nci}$ by substituting Eq. (5.10), instead of $f_{\bar{n}}(v, R)$, into Eq. (5.7) and (5.8). Then we weight them by P and $1-P$ to get the average p_{nc} of the whole system as:

$$\bar{p}_{nc} = P\bar{p}_{nc1} + (1-P)\bar{p}_{nc2}. \tag{5.12}$$

For strategy 1, we have:

$$\bar{p}_{nc1} = 1 - \frac{4(1-p_{01})R_1\mu}{\pi V_T p_{f1}} ln\left[1 + \frac{\pi V_T p_{f1}}{4R_1\mu}\right], \tag{5.13}$$

$$\bar{p}_{nc2} = 1 - \frac{4(1-p_{02})R_2\mu}{\pi(V_{max} - V_T)p_{f2}} ln\left[\frac{4R_2\mu + \pi V_{max}p_{f2}}{4R_2\mu + \pi V_T p_{f2}}\right]. \tag{5.14}$$

For strategy 2, it can be derived that:

$$\bar{p}_{nc1} = \frac{2V_T}{V_{max}} - \frac{4(1-p_{01})R_1\mu}{\pi V_{max}p_{f1}} ln\left[\left(1 + \frac{\pi V_T p_{f1}}{4R_1\mu}\right)\left(\frac{4R_1\mu + \pi V_{max}p_{f1}}{4R_1\mu + \pi V_T p_{f1}}\right)^{\left(\frac{V_{max}}{V_T} - 1\right)}\right], \tag{5.15}$$

$$\bar{p}_{nc2} = 2 - \frac{V_T}{V_{max}} - \frac{4(1-p_{02})(2V_T - V_{max})R_2\mu}{\pi V_{max}V_T p_{f2}} ln\left[\frac{4R_2\mu + \pi V_{max}p_{f2}}{4R_2\mu + \pi V_T p_{f2}}\right]. \tag{5.16}$$

A low $\bar{p}_{nc}$ is a desirable feature since the handoff blocking can be very annoying to a user. Strategy 1 has a lower $\bar{p}_{nc}$ than that of strategy 2. Similarly, we have the average p_b and p_f of the whole system as $\bar{p}_b = Pp_{b1} + (1-P)p_{b2}$ and $\bar{p}_f = Pp_{f1} + (1-P)p_{f2}$.

Imperfect speed estimations

Generally there are two ways to estimate user speed. One is to examine the characteristics of the received signals [11]. In this method, the level crossing rate or the autocovariance of the received signal samples is used. The other is to use cell dwell time estimators [10], based on MS dwell time in previous cells. Since perfect speed estimation is impossible in reality, estimation errors will affect the performance of strategy 1. We may estimate the speed of an MS as $v + \Delta$, where v is its real speed and Δ is an estimation error. If both $v + \Delta$ and v are, say greater than V_T, estimation error does not affect the system performance because the MS is still treated as a high speed one. Otherwise, It will enter the microcell tier and create more handoffs. From Fig. 5. 2, the increase of the average handoff rate of a call over $\Re_1$ due to the inaccurate estimation is:

$$\Delta\Re_{err} = p \int_{V_T-\Delta v}^{V_T} [f_{\bar{n}}(v, R_2) - f_{\bar{n}}(v, R_1)] f_V(v | V_T - \Delta v < v \le V_T)\, dv + p \int_{V_T}^{V_T+\Delta v} [f_{\bar{n}}(v, R_1) - f_{\bar{n}}(v, R_2)] f_V(v | V_T < v \le V_T + \Delta v)\, dv \quad , \qquad (5.17)$$

where p is the probability that the speed estimation error makes an MS being served by a wrong tier, Δv is the absolute value of maximum speed estimation error. From Eq. (5.17) we derive the following result:

$$\Delta\Re_{err} = \frac{\pi p \Delta v^2}{8\mu V_{max}}\left(\frac{1}{R_1} - \frac{1}{R_2}\right). \qquad (5.18)$$

It can be seen that $\Delta\Re_{err}$ is in proportion to Δv^2, but inversely proportion to V_{max}. It is obvious that the larger the V_{max}, the smaller the MS population within the speed range of $V_T - \Delta v \le v \le V_T + \Delta v$.

5. 2 Mobile Membership Simulation

In DS/CDMA cellular systems, an MS belong to one BS when it is not doing soft handoff. However, during soft handoff, the MS communicate to a number of BSs at the same time. Power control is switched among those BSs depending on the membership of the MS. In order to minimize interference contribution, an MS should be power controlled by the BS providing the highest local mean of pilot signal in the forward link. The knowledge of the controlling BS may be required at every time, which determines the MS's membership. During soft handoff, a controlling BS is selected from the active set. Capacity estimation reported in [28] is based on this assumption. The impact of imperfect BS membership reduces the capacity of the system [5].

5. 2. 1 Model of the received pilot

We first look into the scenario of an MS travel between two BSs, BS_0 and BS_1. The BSs are separated by a distance of D meters as shown in Fig. 5. 5. An MS is moving with a speed of v from BS_0 to BS_1 along the BS_0-BS_1 line. Each BS transmits a pilot signal in the forward link for the MS to perform power measurement and demodulation. The received pilot power is given by:

$$y_i(t) = r_i(t)s_i(t) \text{ for } i = 0, 1 . \tag{5.19}$$

where $r_i(t)$ is the power fluctuation due to multipath fading and $s_i(t)$ is the power fluctuation due to shadowing. $r_i(t)$ and $s_i(t)$ are considered to be statistically independent processes. $i = 0, 1$ denotes BS_0 and BS_1, respectively.

$r_i(t)$ is assumed to have a negative exponential first order distribution with unit mean. Its autocovariance function is modeled as:

$$c_{r_i}(\tau) = J_0^2(2\pi f_D\tau) \text{ for } i = 0, 1 , \tag{5.20}$$

where $J_0(\ )$ is the zero-order Bessel's function of the first kind. The Doppler frequency f_D is MS's moving speed v divided by the RF wavelength λ of the received signal.

$s_i(t)$ is usually expressed in terms of its log-version $S_i(t)$ in dB as:

$$s_i(t) = 10^{S_i(t)/10} = e^{\beta S_i(t)} \text{ for } i = 0, 1 , \tag{5.21}$$

where $\beta = \ln(10)/10$. $S_i(t)$ is a Gaussian process which is represented by the sum of a deterministic component and a zero-mean stationary Gaussian process $x_i(t)$ as:

$$S_i(t) = a_i - b_i \log d_i(t) + x_i(t) \text{ for } i = 0, 1, \tag{5.22}$$

where $d_i(t)$ is the MS-BS$_i$ distance at time t. a_i and b_i are the parameters of the mean signal strength of the MS-BS$_i$ link. According to a widely used model, the autocovariance function of $x_i(t)$ is:

$$c_{x_i}(\tau) = \sigma_{x_i}^2 e^{-v|\tau|/\partial} \text{ for } i = 0, 1, \tag{5.23}$$

where ∂ determines how fast the signal sample correlation decays with the spacial distance and σ_{x_i} is the standard deviation. The power measurements are taken by the MS on both links each and every T seconds. Therefore, we have the following sample sequences:

$$y_i(n) = y_i(nT),\ r_i(n) = r_i(nT),\ s_i(n) = s_i(nT) \text{ for } i = 0, 1, \tag{5.24}$$

The decision of membership update is based on the local mean estimation of the pilots from two BSs. The estimations are obtained by low-pass filtering $y_i(n)$ sequences. Because of multipath propagation, the received signal fluctuates rapidly. To measure the actual local mean of the signal, signal strength must be averaged over time by low-pass filter to remove fast fading. Then the selection of the filter parameter will affect the estimation accuracy and the performance measurements discussed later. Given low-pass filter impulse response $h(n)$, its output is the convolution of $y_i(n)$ and $h(n)$:

$$\hat{y}_i(n) = y_i(n) \otimes h(n) \text{ for } i = 0, 1, \tag{5.25}$$

here we assume that:

$$h(n) = \begin{cases} 1/w & \text{for } 0 \le n < w \\ 0 & \text{else} \end{cases}. \tag{5.26}$$

where w is the window length.

5. 2. 2 Performance measurements

The performance measurements considered here are the member switching rate per second and the probability that MS updates its membership to a BS not providing the best pilot strength. The latter is call wrong selection probability P_{ws}. We define the membership as:

$$M(n) = \begin{cases} 0 \text{ for } \dot{y}_0(n) \geq \dot{y}_1(n) \\ 1 \text{ for } \dot{y}_0(n) < \dot{y}_1(n) \end{cases} . \tag{5.27}$$

The membership switching rate is:

$$\omega = \left\lfloor \frac{1}{T} \right\rfloor P\{M(n) \neq M(n-1)\} . \tag{5.28}$$

The wrong selection probability is:

$$P_{ws} = P\left[\frac{\dot{y}_0(n)}{\dot{y}_1(n)} \geq 1, \frac{s_0(n+1)}{s_1(n+1)} < 1\right] + P\left[\frac{\dot{y}_0(n)}{\dot{y}_1(n)} < 1, \frac{s_0(n+1)}{s_1(n+1)} \geq 1\right] . \tag{5.29}$$

The Eq. (5.27) and (5.29) can be easily extended to a three-BS scenario by adding corresponding variables of the third BS, BS_2, such as $\dot{y}_2(n)$. In addition, the analytical expressions of the performance measurements can be found in [12].

5. 2. 3 The simulation algorithm

After parameters initialization, the simulation is conducted under the following steps.

- Step 1: Generate the signal sample sequence $r_i(n)$. To get the sequence with autocovariance function of Eq. (5.20), we sample this autocovariance function by $2\pi f_D T$ to get a sample sequence $c_{r_i}(n)$ for $0 \leq n \leq z$. Notice that $c_{r_i}(n) = c_{r_i}(-n)$. According to the autoregressive modeling theory [46], the parameters of the forming filter is related to the samples by:

$$\begin{bmatrix} c_{r_i}(0) & c_{r_i}(-1) & \dots & c_{r_i}(-z) \\ c_{r_i}(1) & c_{r_i}(0) & \dots & c_{r_i}(n) \\ \dots & \dots & & \dots \\ c_{r_i}(z) & c_{r_i}(z-1) & \dots & c_{r_i}(0) \end{bmatrix} \begin{bmatrix} 1 \\ h_1 \\ \dots \\ h_z \end{bmatrix} = \begin{bmatrix} g^2 \\ 0 \\ \dots \\ 0 \end{bmatrix} . \tag{5.30}$$

Then the sequence generating filter is in the form of:

$$r(n) = g^2 \tilde{r}(n) - \sum_{i=1}^{z} h_i r(n-i) , \tag{5.31}$$

where $\tilde{r}(n)$ is the input sequence of the generating filter and $r(n)$ is the output. Two independent $r(n)$ sequences are generated to represent $r_0(n)$ and $r_1(n)$ respectively.

- Step 2: Generate the signal sample sequence $s_i(n)$. The autocovariance function sampled is Eq. (5.23). Both a_i and b_i are included. Notice that under a two BS scenario, we always have $d_0 + d_1 = D$.
- Step 3: Generate the pilot signal strength sequence $y_i(n)$.
- Step 4: Generate the pilot local mean estimation $\hat{y}_i(n)$ by filtering $y_i(n)$ with $h(n)$.
- Step 5: Obtain the switching rate from Eq. (5.27) and (5.28).
- Step 6: Obtain the wrong selection probability from Eq. (5.29).

5. 3 Numerical Results and Discussions

5. 3. 1 Comparison of the two strategies

Numerical examples are given to illustrate the system performances under both strategies. In this example, $C_1 = 20$, $C_2 = 30$, $R_1 = 380$ m, $R_2 = 1000$ m. The average call holding time is $1/\mu = 100$ seconds. The V_{max} is chosen to be 60 km/h and 120 km/h to show the system performances under different mobility scenarios.

Fig. 5. 3 (a) shows, as expected, strategy 1 has a smaller handoff rate than strategy 2. MS's mobility also affects handoff rate. Larger V_{max} means a higher average MS mobility

and results in more boundary crossings and thus a higher handoff rate. The spectrum partition between the two tiers determines V_T/V_{max} and the $\Delta\Re$. In this example, 19% traffic is carried by macrocells and $\Delta\Re = 10\%$. Handoff rate reduction is at the cost of overall system capacity reduction if the total bandwidth is constant.

The effect of speed estimation error can be seen from an example. Suppose $V_{max} = 60$ km/h, $\Delta v = V_{max}/10$, other system parameters remain the same. From Eq. (5.18) we have $\Delta\Re_{err} = 0.038$ which is much less than $\Re_1$ (<5%) with ideal speed estimation shown in Fig. 5. 3 (a). The reason is that only $2\Delta v/V_{max}$ percent of MSs may be affected by estimation error. Those MS's speed diversity is rather small, between $V_T + \Delta v$ and $V_T - \Delta v$, as shown in Fig. 5. 2. Therefore, strategy 1 is insensitive to speed estimation error.

In a multi-tier system, traffic load has to be distributed into different tiers such that MSs have roughly the same blocking probability regardless of the tier they reside. Fig. 5. 3 (b) shows p_{bi} and p_{fi}. We can not keep $p_{b1} = p_{b2}$ and $p_{f1} = p_{f2}$ with a fixed V_T. Suppose the two tiers are balanced under a light traffic load, say $p_{bi} = p_{fi} = 2 \times 10^{-3}$. When traffic becomes heavy, p_{b2} (p_{f2}) will be much higher than p_{b1} (p_{f1}). An unbalanced system causes high-speed MSs suffering more handoff blocking or dropping. Therefore, we select a V_T to keep the system balanced under heavy traffic. Moreover, under light traffic, MSs in an unbalanced system are hardly blocked and the unfairness is not concerned. Therefore, light traffic scenario is not of our interest.

Fig. 5. 4. compares $\bar{p}_{nc}$ of the two strategies under different V_{max}. It can be seen that the higher V_{max}, the higher the $\bar{p}_{nc}$. But in strategy 1, $\bar{p}_{nc}$ is considerably lower than that in strategy 2. In other words, for the same $\bar{p}_{nc}$, strategy 1 allows higher p_{bi} and p_{fi}. Thus a cell can carry more traffic than one using strategy 2. The reason is that the reduced handoff rate in strategy 1 results in a lower handoff failure probability. If we set $\bar{p}_b = 2\%$ as a reference blocking rate, we find that $\bar{p}_{nc} = 4.5\%$ in strategy 1 and 8% in strategy 2 when $V_{max} = 60$ km/h. If we reduce the $\bar{p}_{nc}$ in strategy 2 to 4.5%, the system capacity, in terms of traffic carrying capability, will decrease from 16.3 Erlangs to 14.9 Erlangs, a 9% reduction in a microcell area. It also can be seen that the $\bar{p}_{nc}$ increases more rapidly with $\bar{p}_b$ if V_{max} is high.

Based on above analysis, it concludes that strategy 1 is more robust to deal with high mobility MSs. Strategy 1 reduces the chance that a high mobility MS crosses cell boundaries and avoids its generating high handoff rate in microcell tier. Traffic balance is considered when we divide MSs into low and high mobility groups. The disadvantage of strategy 1 is due to its overhead on speed estimation. Fortunately, its performance is insensitive to speed estimation errors.

5. 3. 2 Membership simulation results

The following parameters are assumed for user membership simulation.

$$\begin{matrix} D = 2000 \text{ m} \\ T = 0.01 \text{ s} \end{matrix}, \quad \begin{matrix} \partial = 6 \sim 40 \text{ m} \\ v = 13 \text{ m/s} \end{matrix}, \quad \begin{matrix} f = 900 \text{ MHz} \\ \sigma_{x_0} = \sigma_{x_1} = 4 \sim 12 \text{ dB} \end{matrix}, \quad \begin{matrix} a_0 = a_1 = 10 \\ b_0 = b_1 = 33.8 \end{matrix}.$$

Fig. 5. 6. (a) illustrates the autocovariance of the simulated received signals along with the corresponding theoretical autocovariance. The two autocovariances are in very good agreement. Fig. 5. 6. (b) gives a snapshot of the simulation process with the pilot signal received at an MS, its real local mean and estimated local mean. Obviously, the local mean estimation has errors on top of real local mean, whose impact will be examined in the following figures.

In Fig. 5. 7, the various curves refer to the performance metrics at different MS-BS_0 distances. It is shown that the ω is reduced as the window length increases. The longer the window length, the more signal fluctuation is filtered out. Therefore, the local mean estimation becomes more stable resulting in a lower ω. However, there is an optimum window length for reducing the P_{ws}. This can be explained by the fact that the power of the estimation error is dominant at short windows. At longer windows the decrease of the correlation coefficient between the estimation process and the actual shadowing becomes dominant. At the cell boundary, ω can not be greatly reduced unless we accept a non-optimum P_{ws}. As an MS approaches a BS, both ω and P_{ws} are improved. Other BS pilots are hardly higher than that of the approached BS.

In Fig. 5. 8, the effect of shadowing correlation decay length ∂ is shown. It can be seen that the P_{ws} is rather sensitive to the change of $\bar{d}$. At $\bar{d} = 6$ m, even the minimum P_{ws} is quite high. In this case, a high ω (short window length) seems to be unavoidable in order to keep P_{ws} small.

In Fig. 5. 9, the impact of shadowing standard deviation is shown. Better performance in terms of ω and P_{ws} are obtained when $\sigma_{x_i}^2$ is higher. This can be explained by that an increase of $\sigma_{x_i}^2$ corresponds to an increase of the signal power. The errors between simulation results and analytical results become higher for long window size. Because long window results in poor performance, it is not of our interests. The shadowing environment has a rather great impact on P_{ws} but not on ω as we compare Fig. 5. 8. and Fig. 5. 9.

For the case of three BSs being considered, results are given in Fig. 5. 10. Performance degrades dramatically in terms of P_{ws} when an MS approaches the point with equal distance to the three closest BSs. Three relatively similar pilots make the correct BS selection more difficult than the case of two BSs. Away from that point, P_{ws} reduces greatly. However, the ω is not that sensitive to the change of d_0. Only small difference is shown between two d_0 parameters. Therefore, the window length should be selected to give a small P_{ws}, i. e. we use a high switching rate to reduce wrong BS selection and the interference due to that.

The pilot strength from each BS is simulated 400,000 times under each set of given parameters. The order of the forming filters of both $s_i(n)$ and $r_i(n)$ is 50. This makes the 50 consecutive samples of $s_i(n)$ or $r_i(n)$ correlated as required by Eq. (5.23) or (5.20). Beyond that, the correlation among samples can not be kept. This may introduce some error into simulation results when window length is longer than 50. However, the simulation results are in good agreement with analytical ones at the window length of interest.

In most of the cases, the ω is rather stable towards the change of shadowing environment. At the same time, P_{ws} is quite sensitive to environment change as well as the location of an MS. If P_{ws} is expected to be high, to control CDMA mutual interference due to wrong BS selection, switching signalling load has to be increased accordingly. For window length longer than 40, the improvement in reducing ω is negligible while the degradation on P_{ws} becomes significant. Therefore, long windows should be avoided.

Another important observation is that both ω and P_{ws} degrade at very short window length, which can be seen in Fig. 5. 7 ~ Fig. 5. 10. This confirms the idea that fast fading must be removed to select a correct BS.

5. 4 Conclusions

Handoff performance is always an important performance measure of a cellular system. The work presented in this chapter contributes to the control of the handoff process. Pilot local mean estimation can be used in the handoff initiation phase to trigger handoff. Macro/ microcell channel assignment strategy, i. e. strategy 1, can be part of the handoff execution phase. In the future of wireless systems, handoff technology faces the challenges of high mobility users, dissimilar services connections and overlaid cell coverage. Therefore, further investigation needs to be conducted.

We provide the following remarks to conclude this chapter.

- In hierarchical cellular systems, serving fast moving MSs by macrocells can reduce handoff rate, the probability of call not completed and handoff signalling load. Traffic balance should be kept between tiers, based on an appropriate speed threshold, to ensure fair access. The impact of speed estimation error is not significant on the performance of strategy 1.
- The shadowing characteristics affect both MS member switching rate and the wrong BS selection probability. We can not optimize them at the same time. Since the shadowing characteristics may not be uniform within the entire service area, the filter window can be made adaptive in order to have a better compromise between the two metrics.

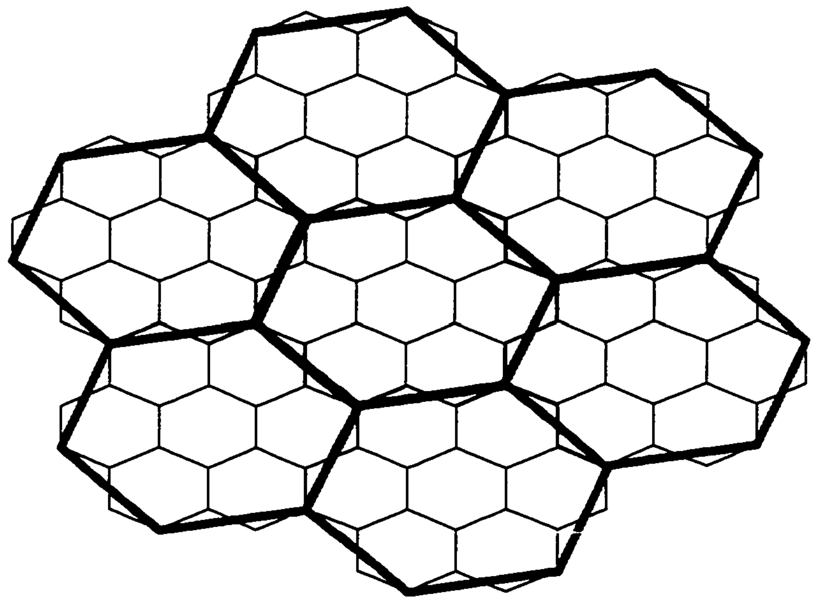


Fig. 5. 1 A two-tier hierarchical wireless cellular system configuration.

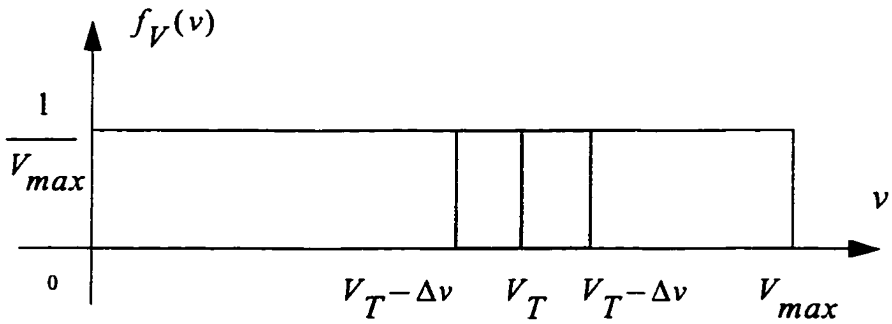


Fig. 5. 2 The impact of imperfect mobile speed estimation on the performance of strategy 1.

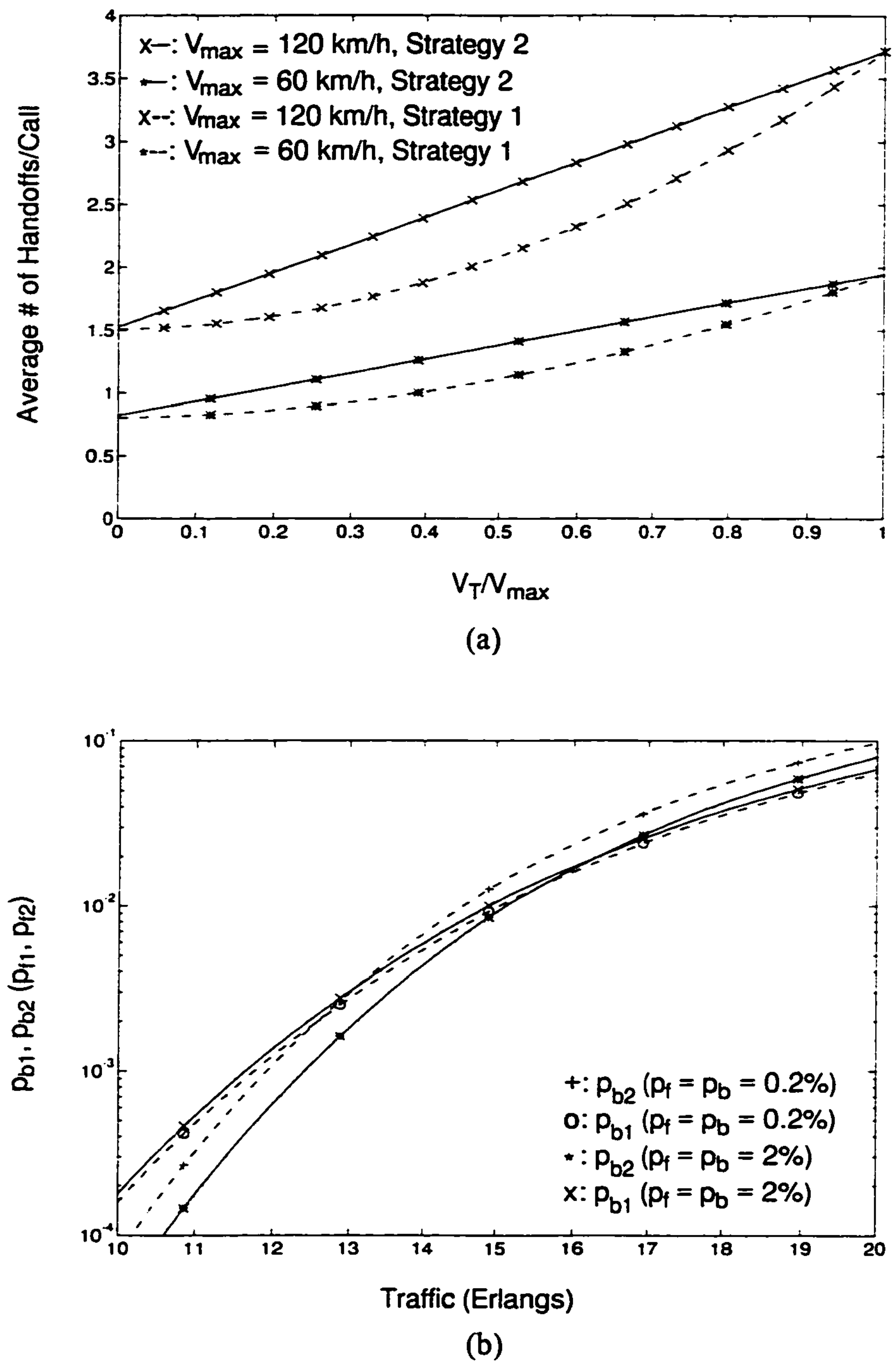


Fig. 5. 3 (a) Average handoff rates of a call for different V_{max} where R_1 = 380 m, R_2 = 1000 m. (b) Call blocking and forced termination probabilities of both tiers. The probabilities at which traffic balanced are given.

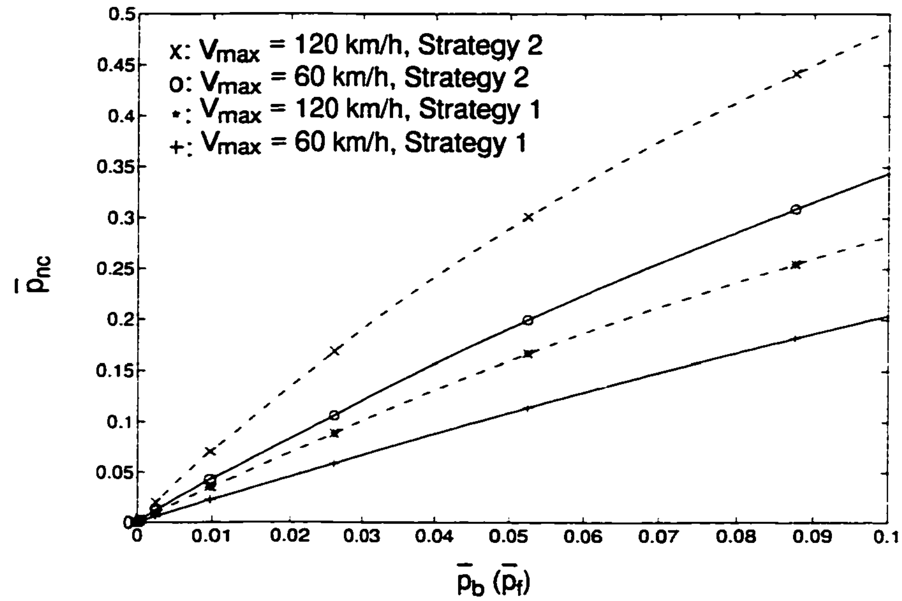


Fig. 5. 4 $\bar{p}_{nc}$ as a function of $\bar{p}_b$ ($\bar{p}_f$). In this case, $C_1 = 30$, $C_2 = 20$, $R_1 = 380$ m, $R_2 = 1000$ m

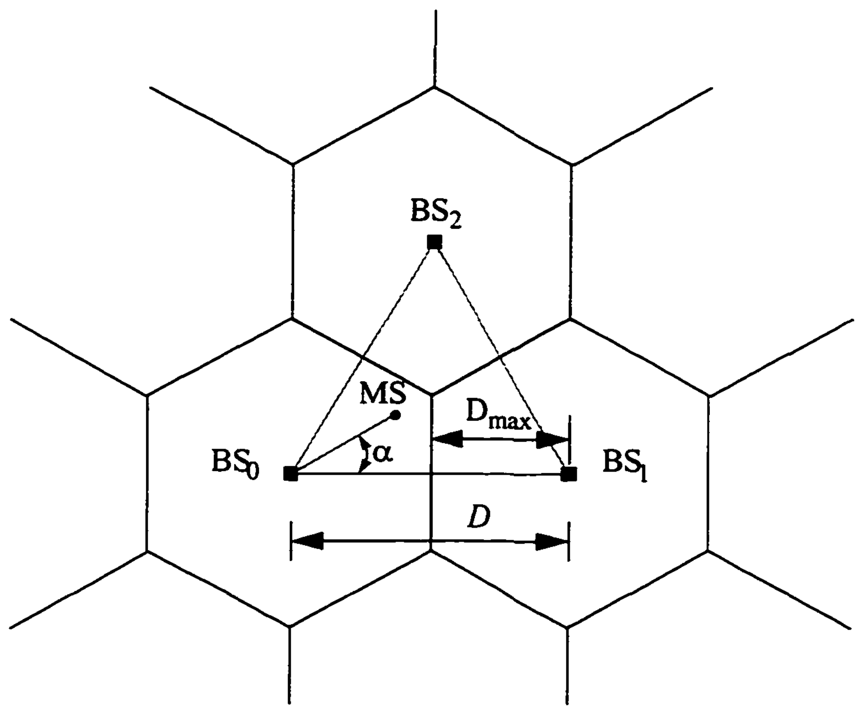


Fig. 5. 5 For the case of two BSs, MS travels from BS_0 to BS_1 ($\alpha = 0$ °). For the case of three BSs, MS travels along a line away from BS_0 and $\alpha = 30$ °.

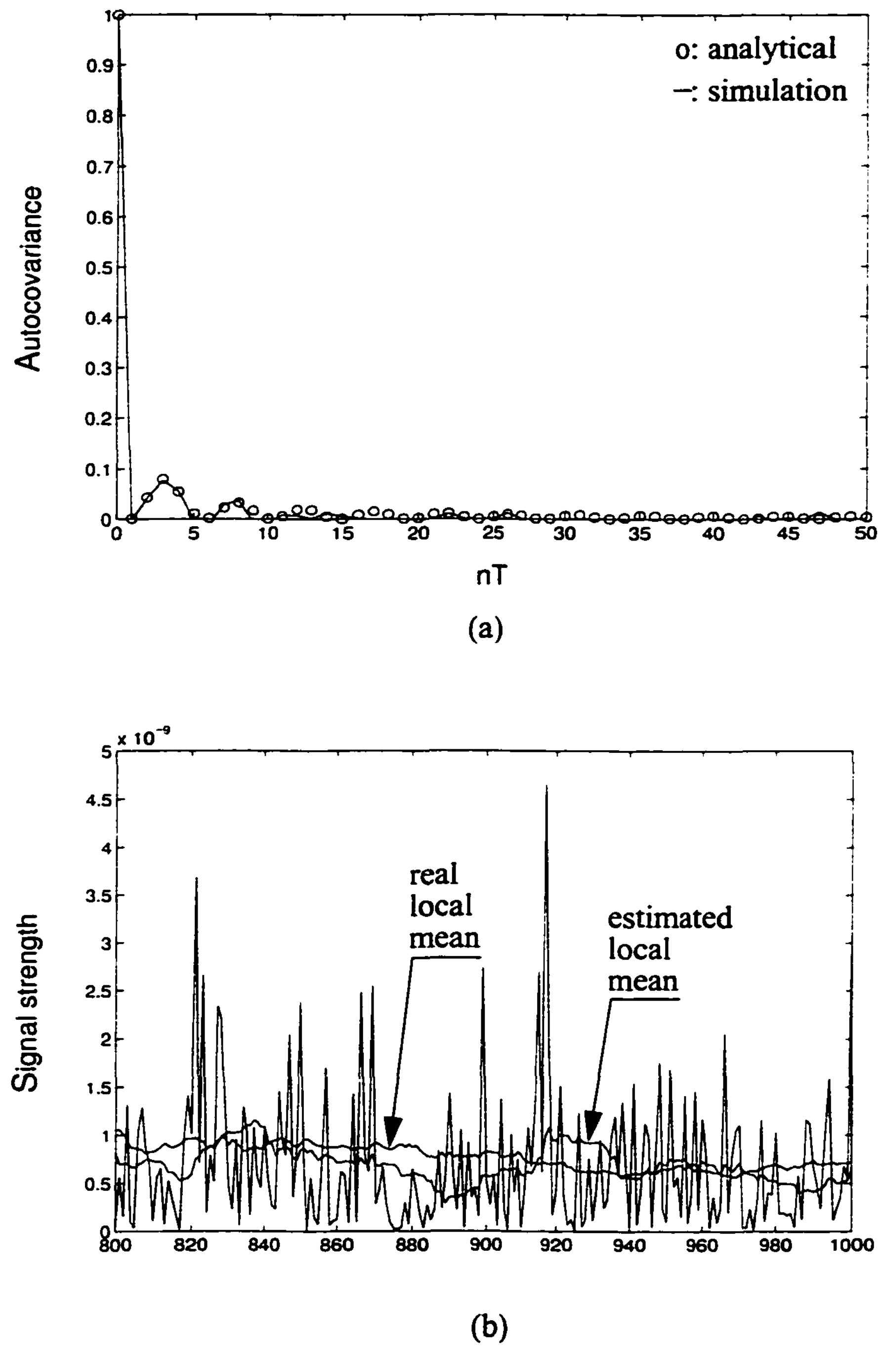
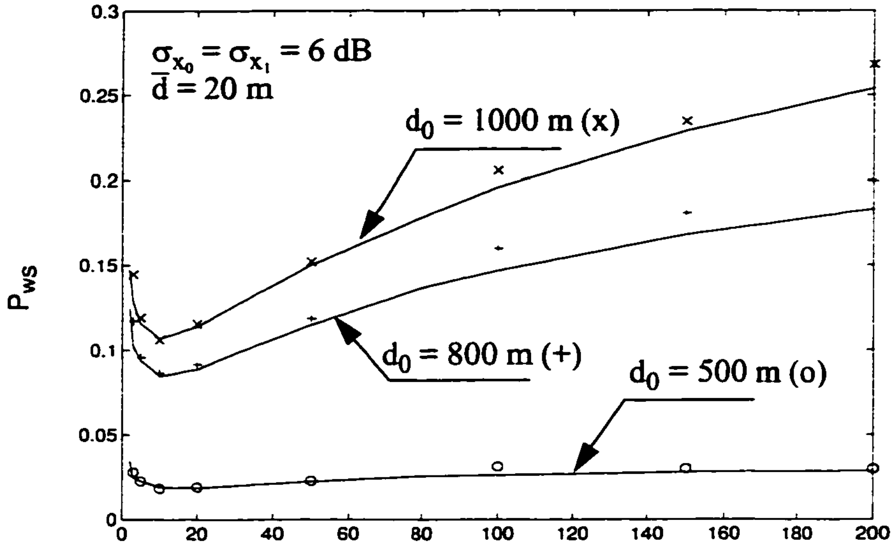
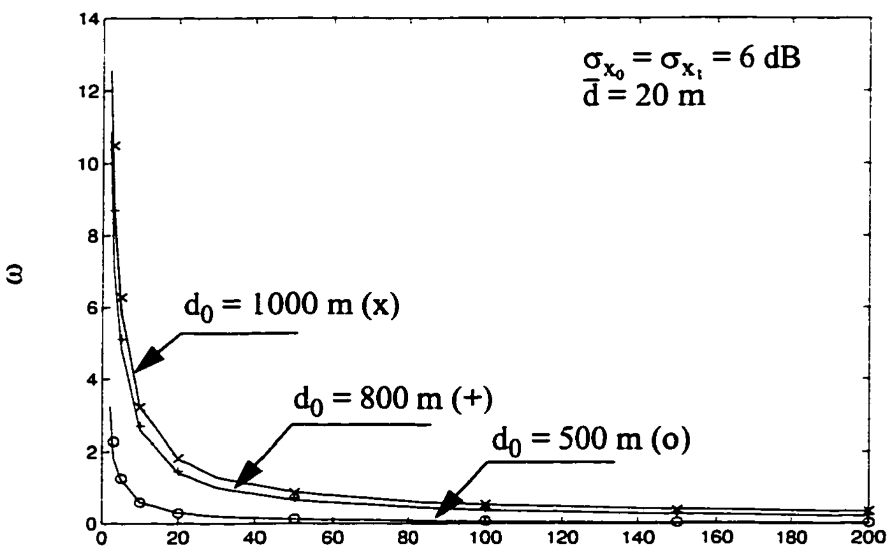


Fig. 5. 6 (a) Simulated autocovariance of the fast fading component of received signal. (b) Simulated received signal and local mean estimations.



(a)



(b)

Fig. 5. 7 (a) Probability of wrong BS selection. (b) Average number of switchings per second. Dots are simulation results.

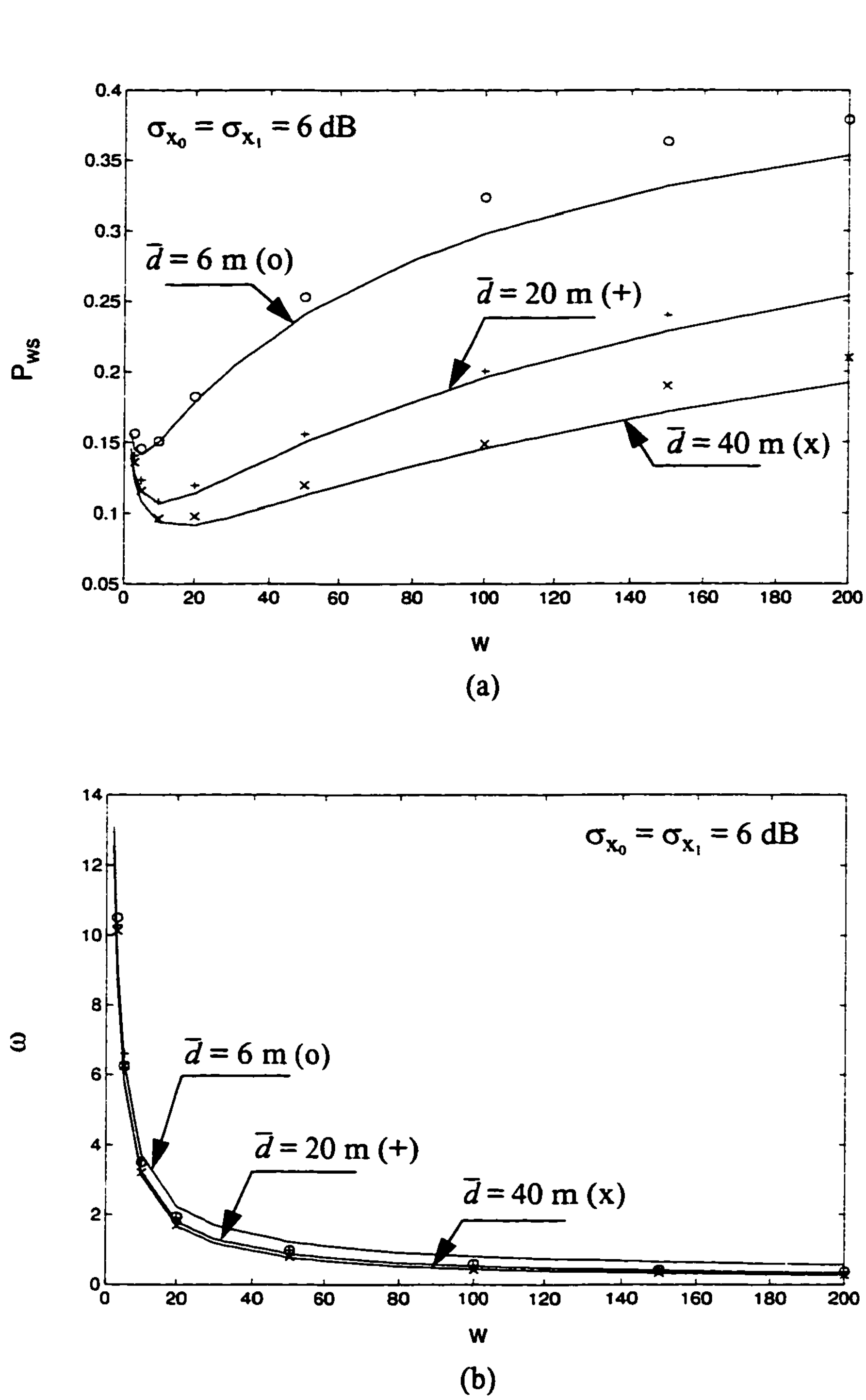
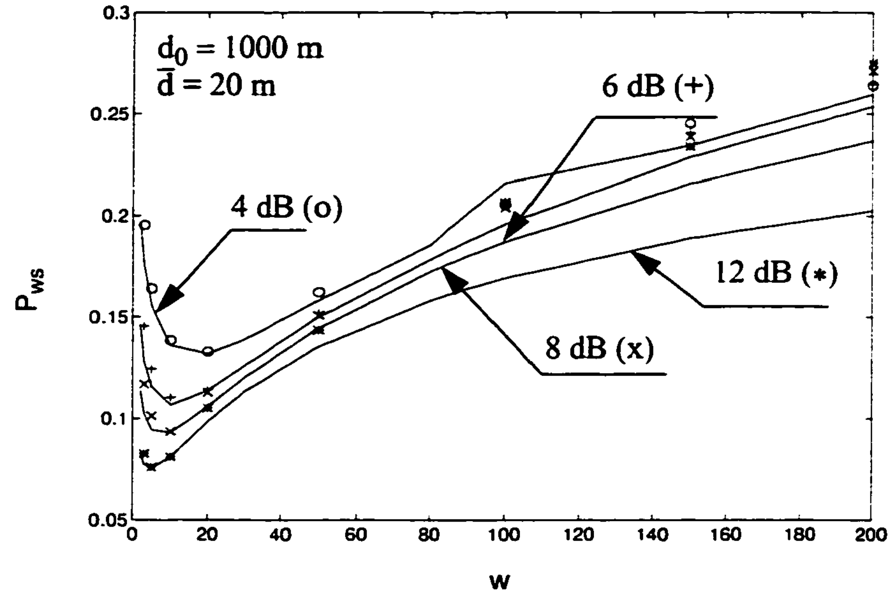
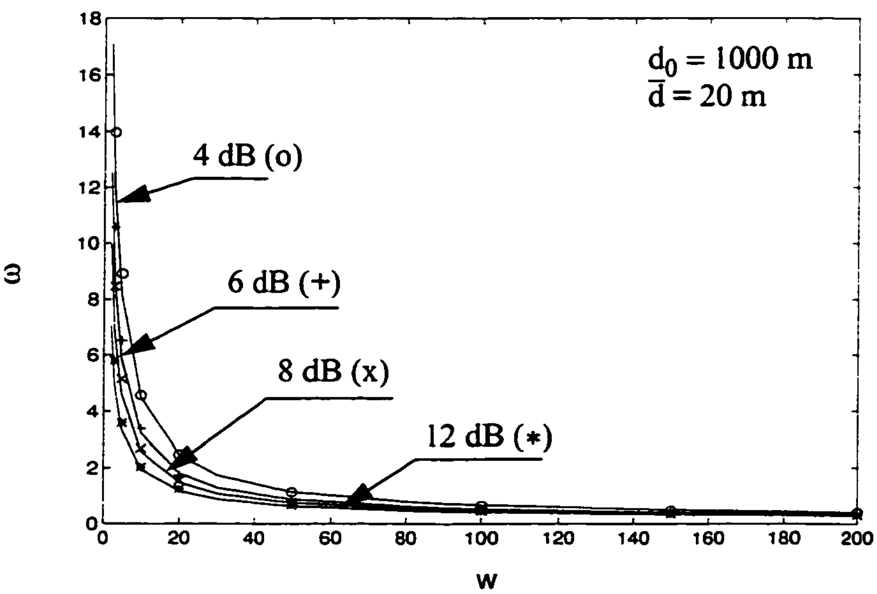


Fig. 5. 8 (a) Probability of wrong BS selection. $d_0 = 1000$ m. (b) Average number of switchings per second. Dots are simulation results.

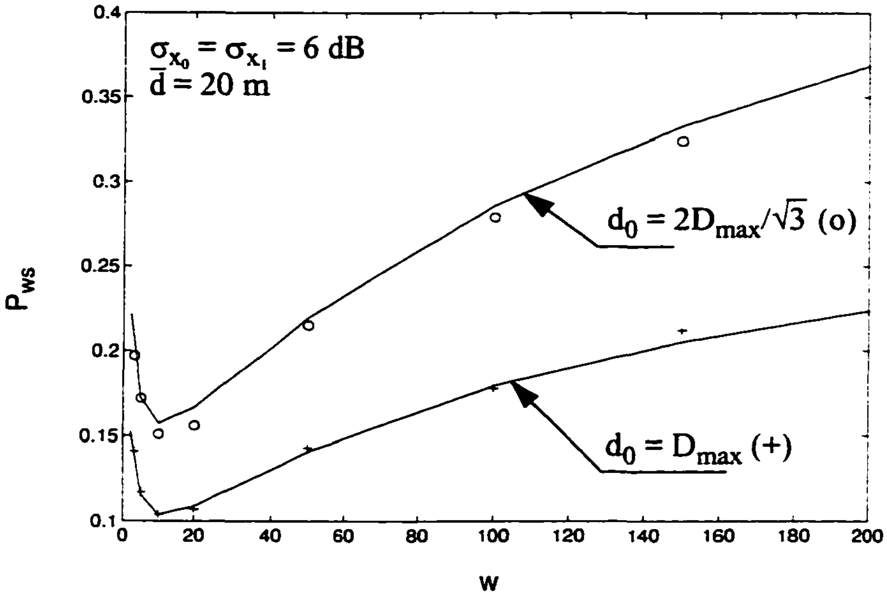


(a)

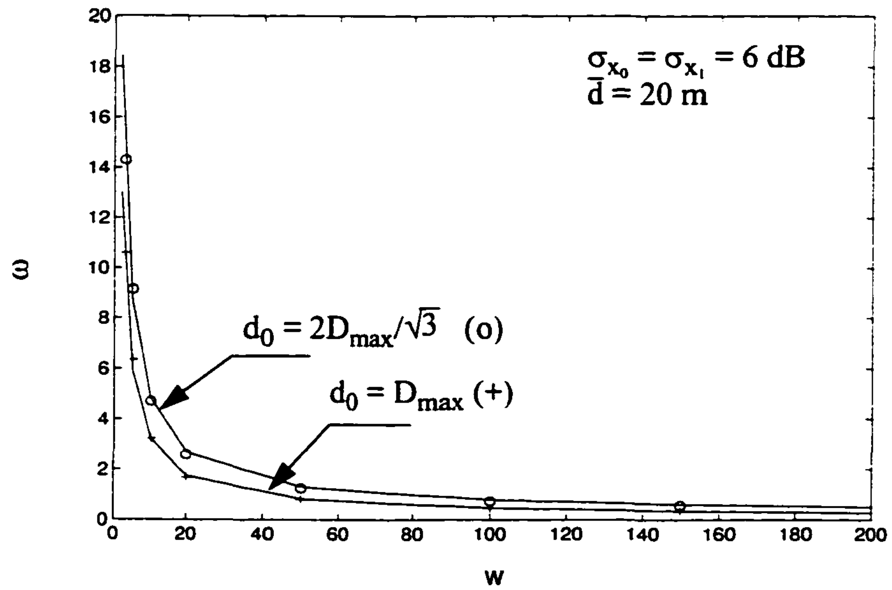


(b)

Fig. 5. 9 (a) Probability of wrong selection. Parameter is the shadowing standard deviation. (b) Average number of switchings per second. Dots are simulation results.



(a)



(b)

Fig. 5. 10 Three base stations ($\alpha = 30$ °). (a) Probability of wrong selection. (b) Average number of switchings per second. Dots are simulation results.

Chapter 6
Conclusions and Future Work

6. 1 Conclusions

In this research, we focused on power control of CDMA systems and multi-service admission control policies to provide a deeper understanding about the problems that future systems will face. Handoff issues are also addressed.

First, based on interference analysis, the reverse link capacity, throughput and delay are studied in a slotted ALOHA DS/CDMA cellular system. We found that the power allocation plays a significant role in maximizing capacity and throughput while keeping the QoS of the services satisfactory. Power control error reduces the capacity. In the case of voice and data service sharing the bandwidth, capacity is much more sensitive to the power control error of voice service. Thus a tight power control, especially for voice service, is necessary. The power control laws are proposed for different cases.

Second, multi-service admission control is more complex than that for single traffic due to service dissimilarity and channel sharing. Reducing blocking probability of important calls and keeping fair access are contradicting goals resulting in that trade-off in admission control becomes important. We have built analytical models to analyze the admission control problems. Several policies are proposed. By access limitation, access enhancement, and soft capacity of CDMA, the relationship of blocking probabilities of all types of calls can be adjusted to fulfill special requirements on blocking probability. In our analysis, handoff calls of one service (or the entire service) are given priority. The reduction in handoff call blocking probability and its cost on the rest are analyzed. The impact of traffic distribution and service dissimilarity, which affect the performance of admission control policies, are examined as well.

Finally, handoff strategies in a hierarchical cellular system are investigated. By assigning macrocell channels to high mobility users and microcell channels to low mobility users, handoff rate and handoff blocking probability decrease in a two tier cellular system. For DS/CDMA cellular systems, user membership simulation is done based on the pilot signal strength estimation. It is found that there is an optimum window length to filter out the fast fading component and minimize the probability of wrong BS selection. This optimum window length is a function of shadowing parameters. Long window reduces power control switching rate. But if it is too long, the probability of wrong BS selection will be high.

6. 2 Suggestions for Future Work

Investigation on the performance of multi-service systems will be important to support the design and deployment of the next generation systems. The following areas are suggested for further research.

- Forward link power control and its performance will become more and more important, since download traffic is expected to be higher than upload in multi-service systems, such as Internet access. This is less addressed since voice traffic on both directions are balanced.
- Future work on multi-service admission control design may consider cells with dynamic channel allocation and policies for variable transmission rate services. Other possible ways of providing prioritization, such as queueing and buffering, need to be examined as well.
- There are still a lot of open issues on multi-tier handoff. Reference [53] provides a good highlight of those problems.

Bibliography

[1] K. Feher, *Wireless digital communications: modulation and spread spectrum applications*, Prentice-Hall Inc., 1995.

[2] P. Andermo, and L. Ewerbring, "A CDMA-based radio access design for UMTS," *IEEE Personal Communications*, vol. 2, no. 1, pp. 48-53, Feb. 1995.

[3] L-R. Hu, and S. S. Rappaport, "Personal communications system using multiple hierarchical Cellular Overlays," *Proc. of ICUPC'94*, pp.397-401, Sept. 1994.

[4] P. Newson, and M. R. Heath, "The capacity of a spread spectrum CDMA system for cellular mobile radio with consideration of system imperfections," *IEEE J. Select. Areas Commun.*, vol. 12, no. 4, pp. 673-684, May 1994.

[5] J. Zou, and V. K. Bhargava, "Reverse link analysis and performance evaluation for DS-CDMA cellular systems," *Proc. of PIMRC'94*, pp. 50-54, Sept. 1994.

[6] W. C. Y. Lee, *Mobile communications design fundamentals*, John Wiley & Sons, Inc., 1993.

[7] M. Frullone, P. Graziaso and G. Riva, "On the optimum allotment of frequency resources in mixed cellular layouts," *IEICE Trans. Fundamental*, vol. E75-A, No. 12, pp.1642-1651, Dec. 1992.

[8] W. Huang, and V. K. Bhargava, "Effects of user mobility on handoff performance in a hierarchical cellular system," *Proc. of CCECE'95*, pp. 551-554, Sept. 1995.

[9] D. Hong, and S. S. Rapaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *IEEE Trans. Veh. Technol.*, vol. 35, pp. 77-92, 1986.

[10] C. W. Sung and W. S. Wong, "User speed estimation and dynamic channel allocation in hierarchical cellular system," *Proc. of VTC'94*, pp.91-95, June 1994.

[11] M. D. Austin, and G.L. Stüber, "Velocity adaptive handoff algorithms for microcellular systems", *IEEE Trans, Veh. Technol.*, vol. 43, pp. 549-561, Aug. 1994.

[12] F. Santucci, W. Huang, and V. K. Bhargava, "Analysis of user membership in CDMA cellular networks," *Proc. of PIMRC'97*, pp. 750-754, Sept. 1997.

[13] C-L. I, L. J. Greenstein and R. D. Gitlin, "A microcell/macrocell cellular architecture for low- and high-mobility wireless users," *IEEE J. Select. Areas Commun.*, vol.11, pp. 885-891, Aug. 1993.

[14] B. Jabbari, "Teletraffic aspects of evolving and next-generation wireless communication networks," *IEEE Personal Communications*, vol. 3, no. 6, pp. 4-9, 1996.

[15] V. O. K. Li, and X. Qui, "Personal communication systems (PCS)," *Proc. of IEEE*, vol. 83, no. 9, pp. 1210-1243, Sept. 1995.

[16] M. E. Anagnostou, and G. C. Manos, "Handover related performance of mobile communication networks," *Proc. of VTC'94*, pp. 111-114, June 1994.

[17] Y. Lin, and S. Mohan, and A. Noerple, "Queueing priority channel assignment strategies for PCS hand-off and initial access," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 704-712, Aug. 1994.

[18] D. E. Everitt, "Traffic engineering of the radio interference for cellular mobile networks," *Proc. of IEEE*, vol. 82, no. 9, pp. 1371-1382, Sept. 1994.

[19] E. Chlebus, and W. Ludwin, "Is handoff traffic really Possionian?," *Proc of ICUPC'95*, pp. 348-353, Nov 1995.

[20] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, and C. E. Wheatley, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 303-312, May 1991.

[21] M. G. Jansen, and R. Prasad, "Capacity, throughput, and delay analysis of a cellular DS CDMA system with imperfect power control and imperfect sectorization," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 57-75, Feb. 1994.

[22] J. Zou, V. K. Bhargava, and Q. Wang, "Reverse link interference modeling and outage analysis for DS-CDMA cellular systems," *Wireless Personal Communications*, vol. 2, pp. 189-215, 1995.

[23] L. F. Fenton, "The sum of log-normal probability distributions in scatter transmission systems," *IRE Trans. Comm. Systems*, vol. C5-8, pp. 57-67, March 1960.

[24] K. I. Kim, "CDMA cellular engineering issues," *IEEE Trans. Veh. Technol.*, vol. 42, no. 3, pp. 345-350, Aug. 1993.

[25] A. Baier, U. Fiebig, W. Granzow, W. Koch, P. Teder, and J. Thielecke, "Design study for a CDMA-based third-generation mobile radio systems," *IEEE J. Select. Areas Commun.*, vol. 12, no. 4, pp. 733-743, May 1994.

[26] V. K. Bhargava, "High rate data transmission in mobile and personal communications," *Proc. PIMRC'94*, pp. 1106-1113, Sept. 1994.

[27] A. Salmasi, and K. S. Gilhousen, "On the system design aspects of code division multiple access (CDMA) applied to digital cellular and personal communications," *Proc. of VTC'91*, pp. 57-62, May 1991.

[28] A. J. Viterbi, A. M. Viterbi, K. S. Gilhousen, and E. Zehavi, "Soft handoff extends CDMA cell coverage and increases reverse link capacity," *IEEE J. Select. Areas Commun.*, vol. 12, no. 8, pp. 1281-1288, Oct. 1994.

[29] S. Sheikh, Y. Yao, and S. Cheng, "Throughput enhancement of direct sequence SS packet radio network by adaptive power control," *IEEE COM*. vol. 42, no. 2/3/4, pp. 884-890, Feb/Mar/Apr. 1994.

[30] W. Huang, and V. K. Bhargava, "Performance evaluation of a DS/CDMA cellular system with voice and data traffic," *Proc. of PIMRC'96*, pp. 588-592, Oct. 1996.

[31] C. I, and K. K. Sabnani, "Variable spreading gain CDMA with adaptive control for true packet switching wireless networkc," *Proc. of ICC'95*, pp. 725-730, June 1995.

[32] A. J. Viterbi, *CDMA: Principle of spread spectrum communication*, Addison-Wesley Publishing Company, 1995.

[33] W. C. Y. Lee, "Overview of cellular CDMA," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 291-302, May 1991.

[34] R. Cameron, and B. D. Woerner, "An analysis of CDMA with imperfect power control", *Proc. of VTC'96*, pp. 977-980, May 1992.

[35] J. Zou, and V. K. Bhargava, "Optimized power allocation for mixed rate traffic in a DS-CDMA cellular system," *Electronic Letters*, No. 22, pp. 1902-1903, Oct. 26, 1995.

[36] K. S. Trivedi, *Probability and statistics with reliability, queuing, and computer science applications*, Prentice-Hall, Inc., 1982.

[37] R. Prasad, M. G. Jansen, and A. Kegel, "Capacity analysis of a cellular direct sequence code division multiple access system with imperfect power control," *ICECE Trans. Commun.*, vol. E76-B, no. 8, pp. 894-905, Aug. 1993.

[38] E. Kudoh, and T. Matsumoto, "Effect of transmitter power control imperfections on capacity in DS/CDMA cellular mobile radios," *Proc. of ICC'92*, pp. 237-242, June 1992.

[39] E. Kudoh, "On the capacity of DS/CDMA cellular mobile radios under imperfect transmitter power control," *ICECE Trans. Commun.*, vol. E76-B, no. 8, pp. 886-893, Aug. 1993.

[40] W. Huang, and V. K. Bhargava, "Capacity and admission control of a slotted DS/CDMA cellular system with voice and data services," submitted to *IEEE Trans. on COM.*

[41] F. Pavlidou, "Mixed media cellular systems," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 848-853, Feb/Mar/April 1994.

[42] P-R. Chang, and B-C. Wang, "Adaptive fuzzy power control for CDMA mobile radio systems," *IEEE Trans. Veh. Technol.*, vol. 45, no. 2, pp. 225-236, May 1996.

[43] C-J. Chang, J-H. Lee, and F-C. Ren, "Design of power control mechanisms with PCM realization for the uplink of a DS-CDMA cellular mobile radio system," *IEEE Trans. Veh. Technol.*, vol. 45, no. 3, pp. 522-530, Aug. 1996.

[44] D. C. Cox, "Wireless personal communications: what is it?," *IEEE Personal Communications*, vol. 2, no. 2, pp. 20-35, April 1995.

[45] J. Rapeli, "UMTS: target, system concept, and standardization in a global framework," *IEEE Personal Communications*, vol. 2, no. 1, pp. 20-28, Feb. 1995.

[46] W. A. Gardner, *Statistical spectral analysis: a nonprobabilistic theory*, Prentice-Hall Inc., New Jersey, 1988.

[47] K. David, T. Kassing, and D. Florack, "ATDMA test objectives," *Proc. of IEEE PIMRC'94*, pp. 691-695, Sept. 1994.

[48] P. Mermelstein, A. Jalali, and H. Leib, "Integrated services on wireless multiple access networks," *Proc. of ICC'93*, pp. 863-867, May 1993.

[49] D. Giancristofaro, M. Ruggieri, and F. Santucci, "Queueing of handover requests in microcell network architecture," *Proc. of VTC'94*, pp. 1846-1849, June 1994.

[50] Y. Lin, A. R. Noerpel, and D. J. Harasty, "The sub-rating channel assignment strategy for PCS hand-offs," *IEEE Trans. Veh. Technol.*, vol. 45, no. 1, pp. 122-130, Feb. 1996.

[51] R. Prasad, M. G. Jansen, and A. Kegel, "Capacity analysis of a cellular direct sequence code division multiple access system with imperfect power control," *ICECE Trans. Commun.*, vol. E76-B, no. 8, pp. 894-905, Aug. 1993.

[52] R. R. Gejji, "Forward-Link-Power control in CDMA cellular systems," *IEEE Trans. Veh. Technol.*, vol. 41, no. 4, pp. 532-536, Nov. 1992.

[53] G.P. Pollini, "Trends in handover design," *IEEE Com. Magazine.*, March 1996.

[54] J-P. Linnartz, *Narrowband land-mobile radio networks*, Artech House, Inc., 1993.

[55] J. Wu and R. Kohno, "A wireless multimedia CDMA system based on transmission power control," *IEEE J. Select. Areas Commun.*, Vol.14, pp.683-691, May, 1996.

[56] A. Annamalai, and V. K. Bhargava, "Some results on FDM/DS-CDMA signalling scheme for wireless communication networks," *Proc. of ICPWC'96*, pp. 116-122, Sept. 1996.

[57] N. Benvenuto, and F. Santucci, "Comparison between least squares path loss estimation and averaging for handover algorithms," *Proc. of VTC'96*, pp. 1326-1330, 1996.

[58] S. Tekinay, and B. Jabbari, "Handover and channel assignment in mobile cellular networks," *IEEE Com. Magazine*, pp. 42-46, Nov. 1991

[59] S. Tekinay, and B. Jabbari, "A measurement-based prioritization scheme for handovers in mobile cellular networks," *IEEE J. Select. Areas Commun.*, Vol.10, no. 8, pp.1343-1350, Oct. 1992.

[60] Y. Lin, A. R. Noerpel, and D. J. Harasty, "The sub-rating channel assignment strategy for PCS hand-offs," *IEEE Trans. Veh. Technol.*, vol. 45, no. 1, pp. 122-130, Feb. 1996.

[61] J. T. Wu, and E. Geraniotis, "Power control in multi-media CDMA networks," *Proc. of VTC'95*, pp. 789-793, 1995.

[62] J. M. G. Linnartz, R. Hekmat, and R. Venema, "Near-Far effects in land mobile random access networks with narrow-band Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 45, no. 1, pp. 122-130, Feb. 1996.

[63] R. Pichna, R. Kerr, Q. Wang, V. K. Bhargava, and I. F. Blake, "CDMA cellular network analysis software," *Final Rep. Ref. No 36-011-2-3506/01-ST*, prepared for Dept. of Commun., CRC, Ottawa, Canada, March 1995.

[64] I. W. Habib, and T. N. Saadawi, "Multimedia traffic characteristics in broadband networks," *IEEE Com. Magazine*, pp. 48-54, July 1992.

[65] W. Yang, and E. Geraniotis, "Admission policies for integrated voice and data traffic in CDMA packet radio networks," *IEEE J. Select. Areas Commun.*, Vol.12, no. 4, pp. 654-664, May 1994.

[66] D. X. Chen, and J. W. Mark, "Performance analysis of output buffered fast packet switches with bursty traffic loading," *Proc. of GLOBECOM'91*, pp. 455-459, 1991.

[67] J. N. Daigle, and J. D. Langford, "Models for analysis of packet voice communications systems," *IEEE J. Select. Areas Commun.*, Vol.4, no. 6, pp. 847-855, May 1986.

[68] A. H. Ang, and W. H. Tang, *Probability concept in engineering planning and design*, John Wiley & Sons Inc., 1984.

[69] S. Ariyavisitakul, and L. F. Chang, "Signal and interference statistics of a CDMA system with feedback power control," *IEEE Trans. Commun.*, Vol.41, no. 11, pp. 1626-1634, Nov. 1993.

Appendix A

Symbols used in Chapter 3:

Ω	Total number of services
m_i	Logarithmic mean of the power of a MS (carrying type-i service)
σ_i	Power control error of a MS (carrying type-i service)
P_i	Received signal power of a MS (carrying type-i service)
$P_{i,j}$	Receiving power of the *j-th* MS (carrying type-i service)
P_t	Total received signal power from MSs within a cell
I	Total intercell interference
n_o	Gaussian noise
v_i	Number of active users of type-i service in a cell
u_i	Total number of users of type-i service in a cell
n_i	Total number of users of type-i service in a cell when C is reached
PG_i	Processing gain of type-i service
$(E_b/N_0)_{thi}$	(E_b/N_0) threshold of type-i service
E_{bi}	Bit energy of type-i service
Ψ_i	Signal to interference ratio threshold of type-i service
α_i	Activity factor of type-i service
G_i	Offered traffic of type-i service
S_i	Throughput of type-i service
D_i	Average packet delay of type-i service
C_i	Capacity that the QoS of type-i service is met
C	Capacity that QoSs of all services are met
$\mathbf{D}$	Delay vector
$\mathbf{D}_{th}$	Delay threshold vector

$\boldsymbol{P}_o$	Outage probability vector
$\boldsymbol{P}_{th}$	Outage probability threshold vector
$\boldsymbol{v}$	Distribution of packets among all services in a slot
$\boldsymbol{u}$	Distribution of MSs among all services in a cell
P_{oi}	Outage probability of a type-i service packet
P_{thi}	Outage probability threshold of type-i service
D_{thi}	Packet delay threshold of type-i service
C_{soft}	Cell soft capacity
G_{mi}	Maximum offered traffic of type-i service
S_{mi}	Maximum throughput of type-i service
$\overline{p_{fi}}$	Average probability a type-i packet being corrupted
r_i	Signal to interference ratio of a type-i packet
$f_{\sigma i}$	Variance of the interference of type-i packets
f_{mi}	Mean of the interference of type-i packets
σ_{Ii}^2	Variance of the interference of total packets
m_{Ii}	Mean of the interference of total packets
$E(\tau)$	Mean of retransmission delay
f_{sect}	Number of sectors per cell
f	Ratio of intercell interference to intracell interference

Appendix B

Given two independent log-normally distributed random variables, x and y with pdf as:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x x}\exp\left[\frac{(\ln x - m_x)^2}{2\sigma_x^2}\right], \tag{B.1}$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_y y}\exp\left[\frac{(\ln y - m_y)^2}{2\sigma_y^2}\right]. \tag{B.2}$$

Let $z = x + y$ be a third random variable and denote:

$$\mu_x = e^{m_x + \sigma_x^2/2},\ \mu_y = e^{m_y + \sigma_y^2/2},\ \text{and}\ \mu = \mu_x + \mu_y. \tag{B.3}$$

$$D_x^2 = e^{2m_x + \sigma_x^2}(e^{\sigma_x^2} - 1),\ D_y^2 = e^{2m_y + \sigma_y^2}(e^{\sigma_y^2} - 1),\ \text{and}\ D^2 = D_x^2 + D_y^2. \tag{B.4}$$

Fenton proved that the pdf of z is approximately log-normal, i.e. in the form of Eq. (B.1) but with the following parameters [23]:

$$\sigma_z^2 = \ln(D^2 + \mu^2) - \ln\mu^2,\ \text{and,}\ m_z = \ln\mu - \sigma_z^2/2. \tag{B.5}$$

By substituting Eq. (B.3) and (B.4) into Eq. (B.5), we have:

$$\sigma_z^2 = \ln\left[\frac{e^{2m_x + 2\sigma_x^2} + 2e^{(m_x + m_y) + (\sigma_x^2 + \sigma_y^2)/2} + e^{2m_y + 2\sigma_y^2}}{e^{2m_x + \sigma_x^2} + 2e^{(m_x + m_y) + (\sigma_x^2 + \sigma_y^2)/2} + e^{2m_y + \sigma_y^2}}\right], \tag{B.6}$$

$$m_z = \ln(e^{m_x + \sigma_x^2/2} + e^{m_y + \sigma_y^2/2}) - \sigma_z^2/2. \tag{B.7}$$